

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 08:54:56 ; Search time 94.52 Seconds
(without alignments)
14.102 Million cell updates/sec

Title: US-09-786-648-3

Perfect score: 62

Sequence: 1 VEVPSQHTDSQ 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	100.0	12	21	AAV87461
2	62	100.0	15	10	AAV87461
3	62	100.0	15	16	AAV87461
4	62	100.0	21	21	AAV87461
5	62	100.0	21	21	AAV87461
6	62	100.0	23	16	AAV87461
7	62	100.0	26	4	AAV87461
8	62	100.0	41	6	AAV87461
9	62	100.0	46	6	AAV87461
10	62	100.0	47	4	AAV87461
11	62	100.0	93	16	AAV87461

12	62	100.0	93	20	AAV87461
13	62	100.0	93	20	AAV87461
14	62	100.0	93	21	AAV87461
15	62	100.0	93	22	AAV87461
16	62	100.0	103	6	AAV87461
17	62	100.0	103	17	AAV87461
18	62	100.0	103	17	AAV87461
19	62	100.0	103	17	AAV87461
20	62	100.0	103	17	AAV87461
21	62	100.0	103	19	AAV87461
22	62	100.0	103	19	AAV87461
23	62	100.0	103	22	AAV87461
24	62	100.0	103	22	AAV87461
25	62	100.0	103	22	AAV87461
26	62	100.0	105	22	AAV87461
27	62	100.0	118	11	AAV87461
28	62	100.0	123	22	AAV87461
29	62	100.0	123	22	AAV87461
30	62	100.0	124	10	AAV87461
31	62	100.0	124	17	AAV87461
32	62	100.0	124	19	AAV87461
33	62	100.0	124	21	AAV87461
34	62	100.0	124	21	AAV87461
35	62	100.0	124	22	AAV87461
36	62	100.0	124	22	AAV87461
37	62	100.0	124	22	AAV87461
38	62	100.0	124	22	AAV87461
39	62	100.0	124	22	AAV87461
40	62	100.0	124	22	AAV87461
41	62	100.0	124	22	AAV87461
42	62	100.0	124	22	AAV87461
43	62	100.0	124	22	AAV87461
44	62	100.0	125	22	AAV87461
45	62	100.0	126	12	AAV87461

ALIGNMENTS

RESULT 1	
ID	AAV87461
AAV87461	standard; peptide; 12 AA.
AAV87461	
03-JUL-2000	(first entry)
XX	Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:3.
XX	Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;
XX	beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;
XX	adjuvant; immune disorder; diarrhoea.
OS	Vibrio cholerae.
OS	Escherichia coli.
PN	MO200014114-A1.
PD	16-MAR-2000.
07-SEP-1999;	99WO-GH02970.
07-SEP-1998;	98GB-0019484.
(UYBR-) UNIV BRISTOL.	
Williams NA, Hirst TR,	
WPI; 2000-256943/22.	
Derivatives of Escherichia coli heat labile enterotoxins useful as	
immunomodulators and for treating diarrhea and which do not bind the	
glycolipid receptor GM-1 -	

XX Disclosure: Page 15; 62pp; English.
PS
CC The invention relates to peptide fragments of the Escherichia coli heat
CC labile enterotoxin (Etx) and its closely related homologue, cholera
CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous
CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are
CC composed of one A subunit and five identical B subunits. The A subunit
CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-
CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)
CC facilitate the entry of subunit A into the host cell via the binding and
CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible
CC for some of the effects of Etx and Ctx, it has been found that certain
CC effects of the toxins, such as immunomodulation, are not mediated
CC through GM-1 binding. The peptides of the invention are fragments of the
CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as
CC normal EtxB and CtxB subunits, except that they do not bind or cross link
CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.
CC They may also be used as an inhibitor for toxin-induced diarrhoea.
CC Therefore, the peptides may be used in the production of a composition
CC for treating, preventing and/or modulating a disease associated with an
CC immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463
CC represent preferred peptides of the invention, AAY87460 being
CC particularly preferred.
XX
SQ Sequence 12 AA:

Query Match 100.0%; Score 62; DB 21; Length 12;
Best Local Similarity 100.0%; Pred. No. 4.6e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHIDSQ 12
|||||
DB 1 vevpgsqhidsq 12

RESULT 2
AAP93498
ID AAP93498 standard; protein; 15 AA.
XX
AC AAP93498;
XX
DT 03-MAY-1990 (first entry)
XX
DE CTP3 epitope of the Cholera toxin B subunit.
XX
KM CTP3 epitope of Cholera toxin B subunit; flagellin fusion protein;
KM vaccine; immunotherapy; ds;
XX
PN WO8910967-A.
XX
PD 16-NOV-1989.
XX
PF 05-MAY-1989; 89WO-US01932.
XX
PR 05-MAY-1988; 88US-0190570.
XX
PA (PRAX-) PRAXIS BIOLOGICS INC.
PA (STRD) LEYLAND STANDFORD JUNIOR UNIV.
XX
PL Marjarian WR, Stocker BAD, Newton SMC;
XX
DR WPI: 1989-356496/48.
DR N-PSDB; AAN92414.
XX
XX
PT New recombinant flagellin gene including sequence - for heterologous
PT epitope, and expressed fusion proteins, useful in vaccines and for prodn.
PT of antibodies.
XX
XX
PS Disclosure: fig.4B; 137pp; English.
CG This sequence corresponds to the CTP3 epitope of the Cholera toxin B

CC subunit. The DNA sequence encoding this ligates to othersynthetic
CC oligonucleotides to form a new recombinant gene. This encodes
CC a flagellin fusion protein which can be used in vaccines for immuno-
CC therapy.
XX
SQ Sequence 15 AA:

Query Match 100.0%; Score 62; DB 10; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.9e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHIDSQ 12
|||||
DB 1 vevpgsqhidsq 12

RESULT 3
AAR85125
ID AAR85125 standard; peptide; 15 AA.
XX
AC AAR85125;
XX
DT 13-JUN-1996 (first entry)
XX
DE Cholera toxin B antigenic peptide fragment CTP3.
XX
DE
XX
KM Conjugate: cholera; B toxin; peptide fragment; microparticulate;
KM inert carrier; modified silica; thyroglobulin; oral vaccine;
KM immunisation; infection; insoluble; digestive tract; antigen;
KM intestines; antibodies; secretory; Iga class.
XX
OS
OS
XX
PN WO9529701-A1.
XX
PD 09-NOV-1995.
XX
PF 02-MAY-1995; 95WO-EP01661.
XX
PR 03-MAY-1994; 94IL-0109519.
XX
PA (YEDA) YEDA RES & DEV CO LTD.
XX
PI Marks RS, Mirelman D, Sela M;
XX
DR WPI: 1995-403805/51.
XX
PT Vaccines for oral immunisation against infecting agents, e.g.
PT cholera - comprise a conjugate of an antigen of an infecting agent
PT covalently bound to micro:particulate inert carrier, e.g. modified
PT aldehyde silica
XX
PS Claim 7; Page 25; 40pp; English.
XX
XX
CC A compsn. comprising a conjugate of an antigenic cholera B toxin
CC peptide fragment (e.g. AAR85126-30 and esp. AAR85125), covalently
CC bound to a microparticulate inert carrier (e.g. modified silica or
CC thyroglobulin) can be used as an oral vaccine for immunisation
CC against cholera infection. The inert carrier is insoluble in the
CC digestive tract, allowing presentation of the antigen in the
CC intestines, where it will elicit antibodies mainly of the
CC secretory Iga class.
XX
SQ Sequence 15 AA:

Query Match 100.0%; Score 62; DB 16; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.9e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHIDSQ 12
|||||

Db 6 vevpgsqhidsq 17

1. *Construcção de J. C. e*

RESULT
AAR76748

ID AAR76748 standard; Protein; 23 AA.
XX AAR76748;
AC
XX
XX 18-MAR-1996 (first entry)
DE Residues 50-64 of cholera toxin B subunit and FimH 224-226.
XX
XX FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW FimH; FimH; receptor binding site; PCR; amplify; ss.
XX
OS Chimeric - Vibrio cholerae.
OS Chimeric - Escherichia coli.
XX
XX
FH Key Location/Qualifiers
FT Misc-difference 1
FT Peptide /note= "Represents FimH residue 224"
FT Peptide 2..4
FT Peptide /note= "Linker peptide"
FT Peptide 5..19
FT Peptide /note= "Cholera toxin B subunit 50-64"
FT Peptide 20..22
FT Misc-difference 23 /note= "Linker peptide"
FT Misc-difference /note= "Represents FimH residue 226"
XX
XX W09520657-A1.
PN 03-AUG-1995.
XX
XX 27-JAN-1995; 95WO-DK00042.
XX
XX 27-JAN-1994; 94US-0187166.
XX
XX (GXRI-) GX BIOSYSTEMS AS.
XX
XX Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX
XX WPI; 1995-275442/36.
DR N-PSDB; AAQ93061.
XX
XX Receptor specific bacterial adhesins - useful for targeting active
PT compounds and microbial cells to locations of receptors
XX
XX Example 3; Page 58; 152pp; English.
PS
XX This sequence is encoded by a fragment of the the plasmid pUPA93
CC which was used in the production of FimH fusion genes comprising
CC the cholera toxin B subunit inserted into the fimH gene. This insert
CC shows the inclusion of the B subunit into the FimH protein at position
CC 224-226. The chimeric genes were then opt. further modified by insertion
CC of the hepatitis B virus surface antigen pre-S2 region into a different
CC position of the FimH adhesin of type 1 fimbriae. Restriction site handles
CC (BglII-sites) were introduced into the fimH gene, and the foreign
CC epitopes are then inserted in-frame. In the selected positions the
CC insertion of the epitopes did not significantly alter the adhesive
CC function of the FimH protein. The expression of the chimeric proteins
CC on the surface of fimbriae on bacterial hosts illustrated the possibility
CC of using bacterial adhesins as general presenters of foreign antigens and
CC epitopes. These chimeric genes may be used in the production of variant
CC FimH adhesins which may be useful for targeting active compounds
CC and microbial cells to locations comprising selected receptors to which
CC the adhesins bind.
XX
XX Sequence 23 AA;
SQ

Query Match 100.0%; Score 62; DB 16; Length 23;
Best Local Similarity 100.0%; Pred. No. 9.6e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VEVPGSQHDSQ 12
|||||

DB 5 vevpgsqhdsq 16
RESULT 7
ID AAP30265 standard; Protein; 26 AA.
XX AAP30265;
AC
XX 21-APR-1992 (first entry)
DE Sequence of amino acids 50-75 of the cholera toxin B1 subunit which
DE carries an Arg at posns. 67 and 73.
XX
XX Cholera vaccine; therapy; E.coli infection; enterotoxin LT.
XX
XX Vibrio cholerae.
XX
XX EP95426-A.
XX
XX 30-NOV-1983.
XX
XX 26-MAY-1983; 83BP-0401052.
XX
XX 26-MAY-1982; 82FR-0009167.
XX
XX (CNRS) CNRS CENT NAT RECH SCI.
XX (INSP) INST PASTEUR.
XX
XX Milhaud G, Raulais D, Rivaille P, Siffert O, Dodin A;
XX Guyongruaz A, Delmas A;
XX
XX WPI; 1983-834645/49.
DR
XX Cholera toxin B, sub-unit polypeptide(s) as vaccines and
PT medicaments - effective against Escherichia coli and Vibrio
PT cholerae infections, are prepd. by solid phase peptide synthesis
XX
XX Claim 7; Page 11; 13pp; French.
PS
XX The inventors claim cholera toxin B1 subunit sequences which carry
CC Arg in posn. 35, 67 or 73 (AAP30264, AAP30265, AAP30600). Tests have
CC shown that Arg in posns. 35, 67 and 73 play an important role in fixing
CC these toxins to cell walls. The peptides are used in the treatment of,
CC and vaccination against cholera infections and animal and human
CC infections due to E. coli (enterotoxin LT). The medicament may be
CC administered by oral, intraperitoneal, sub-cutaneous or intravenous
CC routes. For vaccines, pref. peptides having 15-30 AAs are injected
CC without carriers. Unit dose when used as a medicament is 50-500mg as
CC a vaccine 1-10mg of active cpd.
XX
XX Sequence 26 AA;
SQ

Query Match 100.0%; Score 62; DB 4; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VEVPGSQHDSQ 12
|||||

RESULT 8
ID AAP50439 standard; protein; 41 AA.
XX AAP50439;
AC
XX AAP50439;
XX 01-JAN-1980 (first entry)
DE Network polymer which comprises a series of composite E. coli heat-labile toxin (LT)/heat-stable toxin (ST) polypeptide repeating units.
XX

XX Heat-labile enterotoxin; heat-stable toxin; vaccine; ss.
 KM Synthetic.
 OS
 XX WO8502611-A.
 PN
 XX 20-JUN-1985.
 PD
 XX 12-DEC-1984; 84WO-US02030.
 PF
 XX 12-DEC-1983; 83US-0559469.
 PR
 XX (SCRI-) SCRIpps CLINIC RES.
 PA
 XX Houghten RA;
 PI
 XX WPI; 1985-159230/26.
 DR
 XX
 XX New synthetic polypeptide(s) for part of E. coli enterotoxin(s) -
 PT useful for vaccination of mammals against the enterotoxin(s)
 PS
 XX Claim 8; Page 100; 120pp; English.
 CC The repeating units are bonded together by intramolecular
 CC interpolypeptide cystine bonds formed between oxidized Cys residues
 CC of the repeating units. This polypeptide may be used in the
 CC vaccination of mammals for protection against the enterotoxins. The
 CC composite polypeptide is made by solid phase synthesis or
 CC recombinant DNA techniques. X9 is Met or Lys, X12 is Ala or Glu,
 CC X39 is Asn or Tyr and X46 is Tyr or Asn. See also AAP50439-57.
 CC
 SQ Sequence 41 AA;
 QY
 Db 1 VEVPGSQHDSQ 12
 14 vevpgsqhdsq 25
 Query Match 100.0%; Score 62; DB 6; Length 41;
 Best Local Similarity 100.0%; Pred. No. 0.00018;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 9
 AAP50436
 ID AAP50436 standard; protein: 46 AA.
 XX
 AC AAP50436;
 XX
 DT 01-JAN-1980 (first entry)
 DT
 XX Network polymer which comprises a series of composite E. coli heat-
 DE labile toxin (LT)/heat-stable toxin (ST) polypeptide repeating units.
 DE
 XX Heat-labile enterotoxin; heat-stable toxin; vaccine; ss.
 KM
 XX Synthetic.
 OS
 XX WO8502611-A.
 PN
 XX 20-JUN-1985.
 PD
 XX 12-DEC-1984; 84WO-US02030.
 PF
 XX 12-DEC-1983; 83US-0559469.
 PR
 XX (SCRI-) SCRIpps CLINIC RES.
 PA
 XX Houghten RA;
 PI
 XX WPI; 1985-159230/26.
 DR
 XX

PT New synthetic polypeptide(s) for part of E. coli enterotoxin(s) -
 PT useful for vaccination of mammals against the enterotoxin(s)
 PS
 XX Claim 8; Page 100; 120pp; English.
 CC The repeating units are bonded together by intramolecular
 CC interpolypeptide cystine bonds formed between oxidized Cys residues
 CC of the repeating units. This polypeptide may be used in the
 CC vaccination of mammals for protection against the enterotoxins. The
 CC composite polypeptide is made by solid phase synthesis or
 CC recombinant DNA techniques. X9 is Met or Lys, X12 is Ala or Glu,
 CC X39 is Asn or Tyr and X46 is Tyr or Asn. See also AAP50439-57.
 CC
 SQ Sequence 46 AA;
 QY
 Db 1 VEVPGSQHDSQ 12
 16 vevpgsqhdsq 27
 Query Match 100.0%; Score 62; DB 6; Length 46;
 Best Local Similarity 100.0%; Pred. No. 0.00021;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 10
 AAP30600
 ID AAP30600 standard; protein: 47 AA.
 XX
 AC AAP30600;
 XX
 DT 21-APR-1992 (first entry)
 DT
 XX Sequence of amino acids 350-75 of the cholera toxin B1 subunit which
 DE carries an Arg at posns. 35, 67 and 73.
 DE
 XX Cholera vaccine; therapy; E.coli infection; enterotoxin LT.
 KM
 XX Vibrio cholerae.
 OS
 XX EP95426-A.
 PN
 XX 30-NOV-1983.
 PD
 XX 26-MAY-1983; 83EP-0401052.
 PF
 XX 26-MAY-1982; 82FR-0009167.
 PR
 XX (CNRS) CNRS CENT NAT RECH SCI.
 PA
 XX (INSP) INST PASTEUR.
 PI
 XX Milhaud G, Raulais D, Rivaille P, Siffert O, Dodin A;
 PI Guyongruet A, Delmas A;
 XX WPI; 1983-834645/49.
 DR
 XX Cholera toxin B, sub-unit polypeptide(s) as vaccines and
 PT medicaments - effective against Escherichia coli and Vibrio
 PT cholerae infections, are prepd. by solid phase peptide synthesis
 PT
 XX Claim 8; Page 11; 13pp; French.
 PS
 XX The inventors claim cholera toxin B1 subunit sequences which carry
 CC Arg in posn. 35, 67 or 73 (AAP30264, AAP30265, AAP30600). Tests have
 CC shown that Arg in posns. 35, 67 and 73 play an important role in fixing
 CC these toxins to cell walls. The peptides are used in the treatment of,
 CC and vaccination against, cholera infections and animal and human
 CC infections due to E. coli (enterotoxin LT). The medicament may be
 CC administered by oral, intraperitoneal, sub-cutaneous or intravenous
 CC routes. For vaccines, pref. peptides having 15-30 AAs are injected
 CC without carriers. Unit dose when used as a medicament is 50-500mg as
 CC a vaccine 1-10mg of active cpd.

SQ Sequence 47 AA;

Query Match 100.0%; Score 62; DB 4; Length 47;
 Best Local Similarity 100.0%; Pred. No. 0.00021;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSOHDSQ 12
 |||||
 DB 22 vevpgsqhdsq 33

RESULT 11

AA72545
 ID AAR72545 standard; peptide: 93 AA.

XX AAR72545;

XX 28-NOV-1995 (first entry)

XX ADP-ribosylating toxin (verotoxin-1 B-subunit).

XX ADP-ribosylating toxin; pertussis holotoxin; B-subunit;

XX active site; E. coli heat labile toxin; verotoxin-1;

XX Bordetella pertussis vaccines.

XX Bacteria sp.

XX EP646599-A.

XX 05-APR-1995.

XX 23-AUG-1994; 94EP-0306219.

XX 24-AUG-1993; 93US-0110947.

XX 31-MAY-1994; 94US-0251121.

XX (CONN-) CONNAUGHT LAB LTD.

XX (UYAL-) UNIV ALBERTA.

XX Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;

XX Oomen R, Read RJ, Stein PE;

XX WPI; 1995-132623/18.

XX New modified forms of pertussis holotoxin - developed using

XX crystalline forms of pertussis holotoxin and its complexes with

XX other molecules

XX Disclosure; Fig 5; 54pp; English.

XX AAR72540-R72545 are structurally equivalent B-subunits from three

XX ADP-ribosylating toxins, pertussis holotoxin (PT), E. coli heat

XX labile toxin (LT), and verotoxin-1 (VT). The structural

XX information obt. from these comparisons was used to identify

XX sites which contribute to PT's biological activity. By modifying

XX these sites the claimed PT mutants of the invention were produced,

XX they can be used in the development of vaccines against Bordetella

XX pertussis infection.

XX Sequence 93 AA;

AA41816
 ID AAY41816 standard; peptide: 93 AA.

XX AAY41816;

XX 08-DEC-1999 (first entry)

XX Escherichia coli verotoxin-1 B-subunit.

XX ADP-ribosylating toxin; PT: crystalline pertussis holotoxin; ETA; DT;

XX three-dimensional structure; LT; immunoprotective; infection.

XX Escherichia coli.

XX US5965385-A.

XX 12-OCT-1999.

XX 06-JUN-1995; 95US-0467974.

XX 22-AUG-1994; 94US-0292968.

XX 24-AUG-1993; 93US-0110947.

XX 31-MAY-1994; 94US-0251121.

XX (CONN-) CONNAUGHT LAB LTD.

XX (UYAL-) UNIV ALBERTA.

XX Read RJ, Stein PE, Cockle SA, Loosmore S, Klein MH, Armstrong GD;

XX Hazes B, Oomen RP;

XX WPI; 1999-579908/49.

XX New method for producing modified pertussis holotoxin -

XX Example 3; Fig 5; 41pp; English.

XX A method has been developed of producing a modified pertussis holotoxin,

XX involving analysis of the 3-dimensional form of the crystalline

XX holotoxin. The pertussis holotoxin modification process comprises:

XX (1) identification of at least one amino acid (aa) residue of the

XX holotoxin for modification by analysing the 3-dimensional form of the

XX crystalline holotoxin, in relation to known information of the protein

XX structure and function; (2) effecting mutagenesis (by removing or

XX replacing a nucleotide sequence encoding at least one (aa) of a tox

XX operon; and (3) expressing mutant tox box in a Bordetella organism to

XX produce the modified holotoxin. This method is used for modifying

XX pertussis holotoxin, by studying its 3-dimensional crystalline

XX structure. Modifying the holotoxin, alters its biological properties.

XX By analysing the 3-dimensional crystalline structure of the pertussis

XX holotoxin, functional (aa) which affect biological properties of the

XX pertussis holotoxin can be identified. This can be used to predict (aa)

XX which contribute to the toxicity of the holotoxin to produce

XX immunoprotective, genetically-detoxified analogues of pertussis

XX holotoxin. The present sequence represents an ADP-ribosylating toxin

XX invention.

SQ Sequence 93 AA;

Query Match 100.0%; Score 62; DB 20; Length 93;
 Best Local Similarity 100.0%; Pred. No. 0.00045;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSOHDSQ 12
 |||||
 DB 40 vevpgsqhdsq 51

RESULT 13

AAW95226
 ID AAW95226 standard; peptide: 93 AA.

AC AAW95226;
 XX
 DT 16-MAR-1999 (first entry)
 XX
 DE E. coli heat-labile toxin (LT) beta-subunit sequence.
 XX
 KM Pertussis holotoxin; PT; modified; effector; toxicity; cell binding;
 KM enzyme; ADP-ribosylation; toxin; crystalline; X-ray crystallography;
 KM structural analysis; interacting site; mitogenicity; adjuvant activity;
 KM heat-labile; LT.
 XX
 OS Escherichia coli.
 XX
 PN US5856122-A.
 XX
 PD 05-JAN-1999.
 XX
 PF 22-AUG-1994; 94US-0292968.
 XX
 PR 22-AUG-1994; 94US-0292968.
 PR 24-AUG-1993; 93US-0110947.
 PR 31-MAY-1994; 94US-0251121.
 XX
 PA (UYAL-) UNIV ALBERTA.
 XX
 PI Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;
 PI Oomen RP, Read RJ, Stein PE;
 DR WPI; 1999-105104/09.
 XX
 PT Modifications to e.g. enzymatic activity, mitogenicity and cell
 PT binding of pertussis holotoxin - by identifying interaction sites of
 PT a molecule with crystalline toxin and modifying the identified site
 XX
 PS Example 3; Fig 5; 40pp: English.
 XX
 CC The invention relates to methods of preparing a pertussis holotoxin (PT)
 CC having a modified biological activity. One method comprises identifying
 CC at least 1 site in a PT that interacts with a molecule that is capable of
 CC forming a complex with the holotoxin and which molecule is an effector
 CC molecule which is an adenine nucleotide and which site contributes to
 CC toxicity, cell binding or enzymatic activity of PT. The functional
 CC interacting site(s) are identified by analysing the three dimensional
 CC structure of crystalline PT, determined by X-ray crystallography. The
 CC identified interacting site(s) are modified to alter toxicity, cell
 CC binding or enzyme activity of the PT. The methods can be used to alter a
 CC biological activity such as toxicity, enzymatic activity, mitogenicity,
 CC cell binding and adjuvant activity of the PT. The three-dimensional structure
 CC of PT have functional and/or structural resemblance to other bacterial
 CC toxins such as diphtheria toxin (DT), Pseudomonas exotoxin A (ETA), the
 CC heat-labile toxin of E. coli (LT) and verotoxin-1 (VT). The present
 CC sequence represents the beta-subunit of LT toxin.
 CC
 SQ Sequence 93 AA;

Query Match 100.0%; Score 62; DB 20; Length 93;
 Best Local Similarity 100.0%; Pred. No. 0.00045;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPSQHDSQ 12
 DB 40 vevpsqhdsq 51
 |||||||

RESULT 14
 AAY68365
 ID AAY68365 standard; Peptide; 93 AA.
 AC AAY68365;
 XX
 DT 17-APR-2000 (first entry)
 XX

DE Heat labile toxin B subunit SEQ ID NO:26.
 XX
 KM Pertussis toxin; heat labile toxin; verotoxin; exotoxin A; holotoxin;
 KM diphtheria toxin; ADP-ribosylating toxin; mannose binding protein;
 KM infection; crystal structure; X-ray crystallography; detoxification;
 KM immunogenic.
 XX
 OS Escherichia coli.
 XX
 PN US6018022-A.
 XX
 PD 25-JAN-2000.
 XX
 PF 06-JUN-1995; 95US-0467976.
 XX
 PR 22-AUG-1994; 94US-0292968.
 PR 24-AUG-1993; 93US-0110947.
 PR 31-MAY-1994; 94US-0251121.
 XX
 PA (CONN-) CONNAUGHT LAB LTD.
 PA (UYAL-) UNIV ALBERTA.
 XX
 PI Read RJ, Stein PE, Cockle SA, Loosmore S, Klein MH, Armstrong GD;
 PI Hazes B, Oomen RP;
 DR WPI; 2000-136703/12.
 XX
 PT Crystalline form of isolated pertussis holotoxin useful in studying
 PT proteins which have functional resemblance -
 XX
 PS Example 3; Fig 5; 42pp: English.
 XX
 CC The present invention describes a crystalline form of isolated
 CC pertussis holotoxin, in which the molecules of pertussis toxin have
 CC a three dimensional structure represented in the specification,
 CC complexed with a polysaccharide molecule capable of forming a complex
 CC with the holotoxin. The crystalline form of the pertussis holotoxin
 CC can be used in a comparison with other proteins which have functional
 CC resemblance to pertussis holotoxin with the aim of modifying other
 CC proteins. Identifying the unknown sites of toxicity by comparison
 CC with the three dimensional structure of pertussis holotoxin provides a
 CC technique for detoxification of toxins to produce useful immunogenic
 CC but non-toxic analogues. It can also be used as a primary standard to
 CC measure the quantity, purity or efficacy of less pure compositions
 CC containing pertussis toxin. AAY68340 to AAY68385 represent peptides
 CC used in the exemplification of the present invention.
 CC
 SQ Sequence 93 AA;

Query Match 100.0%; Score 62; DB 21; Length 93;
 Best Local Similarity 100.0%; Pred. No. 0.00045;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPSQHDSQ 12
 DB 40 vevpsqhdsq 51
 |||||||

RESULT 15
 AAB66239
 ID AAB66239 standard; Protein; 93 AA.
 AC AAB66239;
 XX
 DT 03-APR-2001 (first entry)
 XX
 DE E coli verotoxin-1 B subunit SEQ ID NO: 26.
 XX
 KM Pertussis toxin; crystal structure; whooping cough; biological activity;
 KM lymphocytosis-promoting factor; histamine-sensitizing factor;
 KM islet-activating protein.
 XX

OS Escherichia coli.
 XX
 PN US6168928-B1.
 XX
 PD 02-JAN-2001.
 XX
 PF 21-MAY-1998; 98US-0082514.
 XX
 PR 22-AUG-1994; 94US-0292968.
 PR 24-AUG-1993; 93US-0110947.
 PR 31-MAY-1994; 94US-0251121.
 XX
 PA (CONN-) CONNAUGHT LAB LTD.
 XX
 PI Read RJ, Cockle SA, Oomen RP, Loosmore S, Klein MH, Armstrong GD;
 PI Hazes B, Stein PE;
 XX
 DR WPI; 2001-122260/13.
 XX
 PT Modifying pertussis holotoxin to produce detoxified PT analogs,
 PT comprising analyzing crystalline structure of toxin, to identify sites
 PT of toxicity, cell binding or enzyme activity of PT and modifying
 PT identified site -
 XX
 PS Example 3; Fig 5; 41pp; English.
 XX
 CC The present invention provides a method for producing a pertussis toxin
 CC (also designated lymphocytosis-promoting factor, histamine-sensitising
 CC factor and islet activating protein) with a modified biological activity,
 CC involving analysing the crystal structure of the protein to identify
 CC active sites which can then be modified. This may lead to an alteration
 CC in the toxicity, cell binding or enzyme activity of the toxin. This can
 CC be used in the production of immunoprotective analogues of pertussis
 CC toxin. Pertussis toxin is the cause of whooping cough following infection
 CC by Bordetella pertussis.
 XX
 SQ Sequence 93 AA;

Query Match 100.0%; Score 62; DB 22; Length 93;
 Best Local Similarity 100.0%; Pred. No. 0.00045;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 VEYPSQSHTDSQ 12
 |||||
 Db 40 VEYPSQSHTDSQ 51

Search completed: July 3, 2002, 08:54:57
 Job time: 287 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 08:55:38 ; Search time 34.88 Seconds
(without alignments)
8.403 Million cell updates/sec

Title: US-09-786-648-3
Perfect score: 62
Sequence: 1 VEVPGSQHDSQ 12

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCrus.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	100.0	93	2	US-08-292-968-26 Sequence 26, Appl
2	62	100.0	93	2	US-08-467-974-26 Sequence 26, Appl
3	62	100.0	93	2	US-08-467-536-26 Sequence 26, Appl
4	62	100.0	93	3	US-08-467-976-26 Sequence 26, Appl
5	62	100.0	93	4	US-09-082-514-26 Sequence 26, Appl
6	62	100.0	102	3	US-08-952-337-5 Sequence 5, Appl
7	62	100.0	102	3	US-08-952-337-6 Sequence 6, Appl
8	62	100.0	103	2	US-08-472-171-2 Sequence 2, Appl
9	62	100.0	103	2	US-08-894-526-2 Sequence 2, Appl
10	62	100.0	103	4	US-09-013-047-2 Sequence 2, Appl
11	62	100.0	103	4	US-09-374-597-2 Sequence 2, Appl
12	62	100.0	103	5	US-09-191-852-21 Sequence 21, Appl
13	62	100.0	103	5	PCT-US95-13376-21 Sequence 21, Appl
14	62	100.0	123	3	US-08-952-337-1 Sequence 1, Appl
15	62	100.0	123	3	US-08-952-337-2 Sequence 2, Appl
16	62	100.0	124	2	US-08-747-410-2 Sequence 2, Appl
17	62	100.0	371	1	US-08-829-026A-6 Sequence 6, Appl
18	56	90.3	124	1	US-08-449-045C-4 Sequence 4, Appl
19	56	90.3	124	6	US-08-435-605A-12 Sequence 12, Appl
20	56	90.3	124	6	5223610-3 Patent No. 5223610
21	39	62.9	346	2	US-08-602-359A-34 Sequence 34, Appl
22	37	59.7	448	4	US-08-878-989-2 Sequence 2, Appl
23	37	59.7	448	4	US-09-272-796-2 Sequence 2, Appl
24	37	59.7	448	4	US-09-344-700-4 Sequence 4, Appl
25	36	58.1	855	2	US-09-027-337-2 Sequence 2, Appl
26	34	54.8	93	4	US-09-069-023-15 Sequence 15, Appl
27	34	54.8	219	4	US-09-069-023-12 Sequence 12, Appl

28	34	54.8	459	6	5194375-6 Patent No. 5194375
29	34	54.8	775	2	US-08-714-070A-1 Sequence 1, Appl
30	33	53.2	321	3	US-09-039-609-4 Sequence 4, Appl
31	33	53.2	458	3	US-09-039-609-2 Sequence 2, Appl
32	33	53.2	646	4	US-09-625-188-10 Sequence 10, Appl
33	33	53.2	845	1	US-08-416-950-11 Sequence 11, Appl
34	33	53.2	845	2	US-08-469-830-11 Sequence 11, Appl
35	32	51.6	282	1	US-07-672-304-3 Sequence 3, Appl
36	32	51.6	297	1	US-08-011-398B-4 Sequence 4, Appl
37	32	51.6	297	1	US-08-464-051-4 Sequence 4, Appl
38	32	51.6	297	2	US-08-462-498-4 Sequence 4, Appl
39	32	51.6	297	3	US-08-554-385-3 Sequence 3, Appl
40	32	51.6	321	4	US-09-171-461-22 Sequence 22, Appl
41	32	51.6	360	1	US-08-674-612-3 Sequence 3, Appl
42	32	51.6	360	1	US-08-469-421-14 Sequence 14, Appl
43	32	51.6	360	1	US-08-250-975-14 Sequence 14, Appl
44	32	51.6	360	2	US-08-920-296-3 Sequence 3, Appl
45	32	51.6	360	2	US-08-605-002A-14 Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-08-292-968-26
Sequence 26, Application US/08292968
Patent No. 5856122
GENERAL INFORMATION:
APPLICANT: READ, Randy J.
APPLICANT: STREIN, Penelope E.
APPLICANT: COCKLE, Stephen A.
APPLICANT: OOMEN, Raymond P.
APPLICANT: LOOSMORE, Sheena
APPLICANT: KLEIN, Michel H.
APPLICANT: ARMSTRONG, Glen D.
APPLICANT: HAZES, Bart
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/292,968
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/110,947
FILING DATE: 24-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/251,121
FILING DATE: 31-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-388
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-292-968-26

Query Match 100.0%; Score 62; DB 2; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEPGSHIDSQ 12
|||||
Db 40 VEPGSHIDSQ 51

RESULT 2

US-08-467-974-26
; Sequence 26, Application US/08467974
; Patent No. 5965385
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J.
; APPLICANT: STEIN, Penelope E.
; APPLICANT: COCKLE, Stephen A.
; APPLICANT: COOMEN, Raymond P.
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: KLEIN, Michel H.
; APPLICANT: ARMSTRONG, Glen D.
; APPLICANT: HAZES, Bart
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,974
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/467,536
; FILING DATE: 06-JUN-1995
; APPLICATION NUMBER: US 08/292,968
; FILING DATE: 22-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/251,121
; FILING DATE: 31-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/110,947
; FILING DATE: 24-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-454 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-467-974-26

Query Match 100.0%; Score 62; DB 2; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEPGSHIDSQ 12
|||||
Db 40 VEPGSHIDSQ 51

RESULT 3

US-08-467-536-26
; Sequence 26, Application US/08467536
; Patent No. 597304
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J.
; APPLICANT: STEIN, Penelope E.
; APPLICANT: COCKLE, Stephen A.
; APPLICANT: COOMEN, Raymond P.
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: KLEIN, Michel H.
; APPLICANT: ARMSTRONG, Glen D.
; APPLICANT: HAZES, Bart
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,536
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,968
; FILING DATE: 22-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/251,121
; FILING DATE: 31-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/110,947
; FILING DATE: 24-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-455 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-467-536-26

Query Match 100.0%; Score 62; DB 2; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEPGSHIDSQ 12
|||||
Db 40 VEPGSHIDSQ 51

RESULT 4

US-08-467-976-26

```
; Sequence 26, Application US/08467976
; Patent No. 6018022
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J.
; APPLICANT: STEIN, Penelope E.
; APPLICANT: COCKLE, Stephen A.
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: KLEIN, Michel H.
; APPLICANT: ARMSTRONG, Glen D.
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,976
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,968
; FILING DATE: 22-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/251,121
; FILING DATE: 31-MAY-1994
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/110,947
; FILING DATE: 24-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-453 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-467-976-26

Query Match          100.0%; Score 62; DB 3; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 VEVPGSQHDSQ 12
        |||
Db      40 VEVPGSQHDSQ 51
```

```
RESULT      5
US-09-082-514-26
; Sequence 26, Application US/09082514
; Patent No. 6168928
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J.
; APPLICANT: STEIN, Penelope E.
; APPLICANT: COCKLE, Stephen A.
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
```

```
; APPLICANT: ARMSTRONG, Glen D.
; APPLICANT: HAZES, Bart
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/082,514
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,968
; FILING DATE: 24-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-810
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-082-514-26

Query Match          100.0%; Score 62; DB 4; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 VEVPGSQHDSQ 12
        |||
Db      40 VEVPGSQHDSQ 51
```

```
RESULT      6
US-08-952-337-5
; Sequence 5, Application US/08952337
; Patent No. 6019973
; GENERAL INFORMATION:
; APPLICANT: Holmgren, Jan
; APPLICANT: Labens, Michael R.
; TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
; FILE REFERENCE: 3846/0D758
; CURRENT APPLICATION NUMBER: US/08/952,337
; CURRENT FILING DATE: 1998-01-05
; EARLIER APPLICATION NUMBER: PCT/SE96/00570
; EARLIER FILING DATE: 1996-05-02
; EARLIER APPLICATION NUMBER: SE 9501682-0
; EARLIER FILING DATE: 1995-05-05
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Vibrio cholerae
; US-08-952-337-5
```

Query Match 100.0%; Score 62; DB 3; Length 102;
Best Local Similarity 100.0%; Pred. No. 0.00036;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VEPGSOHDSQ 12
DB 49 VEPGSOHDSQ 60

RESULT 7
US-08-952-337-6
; Sequence 6, Application US/08952337
; Patent No. 601973
; GENERAL INFORMATION:
; APPLICANT: Holmgren, Jan
; APPLICANT: Lebens, Michael R.
; TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
; TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS
; FILE REFERENCE: 3846/00758
; CURRENT APPLICATION NUMBER: US/08/952,337
; CURRENT FILING DATE: 1998-01-05
; EARLIER APPLICATION NUMBER: PCT/SE96/00570
; EARLIER FILING DATE: 1996-05-02
; EARLIER APPLICATION NUMBER: SE 9501682-0
; EARLIER FILING DATE: 1995-05-05
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-952-337-6

Query Match 100.0%; Score 62; DB 3; Length 102;
Best Local Similarity 100.0%; Pred. No. 0.00036;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VEPGSOHDSQ 12
DB 49 VEPGSOHDSQ 60

RESULT 8
US-08-472-171-2
; Sequence 2, Application US/08472171
; Patent No. 5932714
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena M.
; APPLICANT: Yacoub, Reza K.
; APPLICANT: Zealey, Gavin R.
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Expression of Gene Products From
; TITLE OF INVENTION: Genetically Manipulated Strains of Bordetella
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 330 University Avenue, Suite 701
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,171
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/393,334

; FILING DATE: 23-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 416-595-1155
; TELEFAX: 416-595-1163
; TELEX: 065-24567 Simbas
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-472-171-2

Query Match 100.0%; Score 62; DB 2; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.00036;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VEPGSOHDSQ 12
DB 50 VEPGSOHDSQ 61

RESULT 9
US-08-894-526-2
; Sequence 2, Application US/08894526
; Patent No. 5942418
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena M
; APPLICANT: Yacoub, Reza K
; APPLICANT: Zealey, Gavin R
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: EXPRESSION OF GENE PRODUCTS FROM
; TITLE OF INVENTION: GENETICALLY MANIPULATED STRAINS OF BORDETELLA
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,526
; FILING DATE: 01-DEC-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-724 MIS-JB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-894-526-2

Query Match 100.0%; Score 62; DB 2; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.00036;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSOHDSQ 12

Db 50 VEVPGSOHDSQ 61

RESULT 10

US-09-013-047-2

; Sequence 2, Application US/09013047
; Patent No. 5998168

; GENERAL INFORMATION:

; APPLICANT: Loosmore, Sheena M.

; APPLICANT: Yacoub, Reza K.

; APPLICANT: Zealey, Gavin R.

; APPLICANT: Klein, Michel H.

; TITLE OF INVENTION: Expression Of Gene Products From

; TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella

; NUMBER OF SEQUENCES: 56

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: 330 University Avenue, 6th Floor

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/013,047

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/472,171

; FILING DATE: 07-JUN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/393,334

; FILING DATE: 23-FEB-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Stewart, Michael I.

; REGISTRATION NUMBER: 24,973

; REFERENCE/DOCKET NUMBER: 1038-507 MIS:vg

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 416-595-1155

; TELEFAX: 416-595-1163

; TELEX: 065-24567 Simbas

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 103 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-09-013-047-2

Query Match 100.0%; Score 62; DB 2; Length 103;

Best Local Similarity 100.0%; Pred. No. 0.00036;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSOHDSQ 12

Db 50 VEVPGSOHDSQ 61

RESULT 11

US-09-374-597-2

; Sequence 2, Application US/09374597
; Patent No. 6140082

; GENERAL INFORMATION:

; APPLICANT: Loosmore, Sheena M.

; APPLICANT: Yacoub, Reza K.

; APPLICANT: Zealey, Gavin R.

; APPLICANT: Klein, Michel H.

; TITLE OF INVENTION: Expression Of Gene Products From

; TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella

; NUMBER OF SEQUENCES: 56

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: 330 University Avenue, 6th Floor

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/374,597

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/393,334

; FILING DATE: FEBRUARY 23, 1995

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Stewart, Michael I

; REGISTRATION NUMBER: 24973

; REFERENCE/DOCKET NUMBER: 1038-964

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 416-595-1155

; TELEFAX: 416-595-1163

; TELEX: 065-24567 Simbas

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 103 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-09-374-597-2

Query Match 100.0%; Score 62; DB 4; Length 103;

Best Local Similarity 100.0%; Pred. No. 0.00036;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSOHDSQ 12

Db 50 VEVPGSOHDSQ 61

RESULT 12

US-09-191-852-21

; Sequence 21, Application US/09191852

; Patent No. 6194560

; GENERAL INFORMATION:

; APPLICANT: Charles J. Arntzen, Hugh S. Mason, and Tariq A. Haq

; TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fulbright & Jaworski L.L.P.

; STREET: 1301 McKinney, Suite 5100

; CITY: Houston

; STATE: TX

; COUNTRY: USA

; ZIP: 77010

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/191,852
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13376
FILING DATE: 24-OCT-1995
APPLICATION NUMBER: 08/817,906
FILING DATE: 04-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Fox, David L.
REGISTRATION NUMBER: 40,612
REFERENCE/DOCKET NUMBER: P015900S1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-651-5151
TELEFAX: 713-651-5246
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-09-191-852-21

Query Match 100.0%; Score 62; DB 4; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.00036;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12
|||||
DB 50 VEVPGSQHDSQ 61

RESULT 13
PCT-US95-13376-21
Sequence 21, Application PC/TUS9513376
GENERAL INFORMATION:
APPLICANT: The Texas A&M University System
APPLICANT: 310 Wisenbaker
APPLICANT: College Station, Texas 77843-3369
TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pravel, Hewitt, Kimball & Krieger
STREET: 1177 West Loop South, 10th Floor
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77027-9095
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13376
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/328,716
FILING DATE: 24-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Jones, John W.
REGISTRATION NUMBER: 31,380
REFERENCE/DOCKET NUMBER: 36170/3P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-850-0909
TELEFAX: 713-850-0165
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear
PCT-US95-13376-21

Query Match 100.0%; Score 62; DB 5; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.00036;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12
|||||
DB 50 VEVPGSQHDSQ 61

RESULT 14
US-08-952-337-1
Sequence 1, Application US/08952337
Patent No. 6019973
GENERAL INFORMATION:
APPLICANT: Holmgren, Jan
APPLICANT: Lebens, Michael R.
TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS
FILE REFERENCE: 3846/0D758
CURRENT APPLICATION NUMBER: US/08/952,337
CURRENT FILING DATE: 1998-01-05
EARLIER APPLICATION NUMBER: PCT/SE96/00570
EARLIER FILING DATE: 1996-05-02
EARLIER APPLICATION NUMBER: SE 9501682-0
EARLIER FILING DATE: 1995-05-05
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 123
TYPE: PRT
ORGANISM: Vibrio cholerae
US-08-952-337-1

Query Match 100.0%; Score 62; DB 3; Length 123;
Best Local Similarity 100.0%; Pred. No. 0.00044;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12
|||||
DB 70 VEVPGSQHDSQ 81

RESULT 15
US-08-952-337-2
Sequence 2, Application US/08952337
Patent No. 6019973
GENERAL INFORMATION:
APPLICANT: Holmgren, Jan
APPLICANT: Lebens, Michael R.
TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS
FILE REFERENCE: 3846/0D758
CURRENT APPLICATION NUMBER: US/08/952,337
CURRENT FILING DATE: 1998-01-05
EARLIER APPLICATION NUMBER: PCT/SE96/00570
EARLIER FILING DATE: 1996-05-02
EARLIER FILING DATE: 1995-05-05
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 123
TYPE: PRT
ORGANISM: Escherichia coli
US-08-952-337-2

Query Match 100.0%; Score 62; DB 3; Length 123;
Best Local Similarity 100.0%; Pred. No. 0.00044;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy	1	VEVPGSOHDSQ	12
Db	70	VEVPGSOHDSQ	81

Search completed: July 3, 2002, 08:55:39
Job time: 214 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 09:03:20 ; Search time 401.04 seconds
(without alignments)
10.532 Million cell updates/sec

Title: US-09-786-648-3
Perfect score: 62
Sequence: 1 VEPGSHINDSQ 12

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues
Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending_Patents_AA_Main:*
1: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
22: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
24: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
25: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
26: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	62	100.0	12	21	US-09-786-648-3
2	62	100.0	15	11	US-08-732-371-1
3	62	100.0	15	11	US-08-732-371A-1
4	62	100.0	21	21	US-09-786-648-4
5	62	100.0	21	21	US-09-786-648-5
6	62	100.0	93	5	US-08-110-947-10
7	62	100.0	93	5	US-08-110-947A-26

8	62	100.0	93	6	US-08-251-121-26	Sequence 26, Appl
9	62	100.0	103	7	US-08-393-334-2	Sequence 2, Appl
10	62	100.0	103	11	US-08-782-832-15	Sequence 15, Appl
11	62	100.0	103	11	US-08-817-906-21	Sequence 21, Appl
12	62	100.0	103	22	US-09-836-433-14	Sequence 14, Appl
13	62	100.0	116	22	US-09-836-433-20	Sequence 20, Appl
14	62	100.0	119	22	US-09-836-433-22	Sequence 22, Appl
15	62	100.0	124	1	PCT-US99-30747-55	Sequence 55, Appl
16	62	100.0	124	1	PCT-US99-30747-57	Sequence 57, Appl
17	62	100.0	124	18	US-09-470-124-55	Sequence 55, Appl
18	62	100.0	124	18	US-09-470-124-57	Sequence 57, Appl
19	62	100.0	138	13	US-08-914-479-2	Sequence 2, Appl
20	62	100.0	138	13	US-08-914-479A-2	Sequence 2, Appl
21	62	100.0	313	21	US-09-756-983-15	Sequence 15, Appl
22	62	100.0	351	21	US-09-756-983-18	Sequence 18, Appl
23	62	100.0	364	21	US-09-756-983-22	Sequence 22, Appl
24	62	100.0	371	5	US-08-150-305A-3	Sequence 3, Appl
25	62	100.0	371	11	US-08-784-218-6	Sequence 6, Appl
26	62	100.0	371	12	US-08-829-026-5	Sequence 5, Appl
27	62	100.0	124	21	US-09-760-234-7	Sequence 7, Appl
28	56	90.3	382	1	PCT-US01-08582-3	Sequence 3, Appl
29	56	90.3	382	1	PCT-US01-08582-4	Sequence 4, Appl
30	56	90.3	461	14	US-09-051-315-2	Sequence 2, Appl
31	56	90.3	461	18	US-09-423-493-2	Sequence 2, Appl
32	56	90.3	750	18	US-09-402-100-2	Sequence 2, Appl
33	56	90.3	1338	18	US-09-402-100-4	Sequence 4, Appl
34	44	71.0	484	16	US-09-252-991A-29252	Sequence 29252, A
35	44	67.7	101	15	US-09-134-000-3773	Sequence 3773, Ap
36	42	67.7	631	26	US-60-324-109-19962	Sequence 1962, A
37	42	67.7	2339	21	US-09-733-089-18876	Sequence 18876, A
38	42	67.7	2339	22	US-09-816-660-18876	Sequence 18876, A
39	40	64.5	769	16	US-09-252-991A-17737	Sequence 17737, A
40	39	62.9	7	21	US-09-786-648-2	Sequence 2, Appl
41	39	62.9	91	19	US-09-595-298A-470	Sequence 470, App
42	39	62.9	114	20	US-09-620-394A-4327	Sequence 4327, Ap
43	39	62.9	115	19	US-09-595-298A-468	Sequence 468, App
44	39	62.9	131	26	US-60-171-481-1502	Sequence 1502, Ap
45	39	62.9	318	1	PCT-US99-17130-318	Sequence 318, App

ALIGNMENTS

RESULT 1
US-09-786-648-3
; Sequence 3, Application US/09786648
; GENERAL INFORMATION:
; APPLICANT: Williams, Neil Andrew
; TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vacci
; FILE REFERENCE: 7438
; CURRENT APPLICATION NUMBER: US/09/786, 648
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: PCT/GB99/02970
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: MS DOS
; SEQ ID NO 3
; LENGTH: 12
; TYPE: PRT
; ORGANISM: E. coli
; FEATURE:
; LOCATION: 50...61
; OTHER INFORMATION: Isolated or synthetic EtXB beta2-alpha2 loop fragment derivabl
; OTHER INFORMATION: human variant E. coli
US-09-786-648-3

Query Match 100.0%; Score 62; DB 21; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY      1 VEVPGSQHIDSQ 12
          |||||
Db      1 VEVPGSQHIDSQ 12
```

RESULT 2

```

1  US-08-732-371-1
2  Sequence 1, Application US/08732371
3  GENERAL INFORMATION:
4  APPLICANT: MIRELMAN, David
5  APPLICANT: MARKS, Robert S.
6  APPLICANT: SELA, Michael
7  TITLE OF INVENTION: VACCINES FOR ORAL IMMUNIZATION AGAINST
8  TITLE OF INVENTION: INFECTING AGENTS
9  NUMBER OF SEQUENCES: 6
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: BROWDY AND NEWMARK
12 STREET: 419 Seventh Street, N.W., Suite 300
13 CITY: Washington
14 STATE: D.C.
15 COUNTRY: USA
16 ZIP: 20004
17 COMPUTER READABLE FORM:
18 MEDIUM TYPE: Floppy disk
19 COMPUTER: IBM PC compatible
20 OPERATING SYSTEM: PC-DOS/MS-DOS
21 SOFTWARE: PatentIn Release #1.0, Version #1.30
22 CURRENT APPLICATION DATA:
23 APPLICATION NUMBER: US/08/732,371
24 FILING DATE: 09-JAN-1997
25 CLASSIFICATION: 424
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER: IL 109519
28 FILING DATE: 03-MAY-1994
29 ATTORNEY/AGENT INFORMATION:
30 NAME: YUN, Allen C.
31 REGISTRATION NUMBER: 37,971
32 REFERENCE/DOCKET NUMBER: MIRELMAN-3
33 TELECOMMUNICATION INFORMATION:
34 TELEPHONE: 202-628-5197
35 TELEFAX: 202-37-3528
36 INFORMATION FOR SEQ ID NO: 1:
37 SEQUENCE CHARACTERISTICS:
38 LENGTH: 15 amino acids
39 TYPE: amino acid
40 TOPOLOGY: linear
41 MOLECULE TYPE: protein
42 US-08-732-371-1

```

Query Match	100.0%	Score 62	DB 11	Length 15
Best Local Similarity	100.0%	Pred. No. 0.00053		
Matches 12: Conservative	0	Mismatches 0	Indels 0	Gaps 0
Qy	1	VEVPGSCHIDSO	12	
Db	1	VEVPGSCHIDSO	12	

RESULT 3

US-08-732-371A-1
Sequence 1, Application US/08732371A
GENERAL INFORMATION:
APPLICANT: MIRELMAN, David
APPLICANT: MARKS, Robert S.
APPLICANT: SELA, Michael
TITLE OF INVENTION: VACCINES FOR ORAL IMMUNIZATION AGAINST
TITLE OF INVENTION: INFECTING AGENTS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington

```

1 STATE: D.C.
2
3 COUNTRY: USA
4
5 ZIP: 20004
6
7 COMPUTER READABLE FORM:
8
9 MEDIUM TYPE: Floppy disk
10
11 COMPUTER: IBM PC compatible
12
13 OPERATING SYSTEM: PC-DOS/MS-DOS
14
15 SOFTWARE: Patentin Release #1.0, Version #1.30
16
17 CURRENT APPLICATION DATA:
18
19 APPLICATION NUMBER: US/08/732,371A
20
21 FILING DATE: 09-JAN-1997
22
23 CLASSIFICATION: 424
24
25 PRIOR APPLICATION DATA:
26
27 APPLICATION NUMBER: IL 109519
28
29 FILING DATE: 03-MAY-1994
30
31 ATTORNEY/AGENT INFORMATION:
32
33 NAME: YUN, Allen C.
34
35 REGISTRATION NUMBER: 37,971
36
37 REFERENCE/DOCKET NUMBER: MIRELMAN-3
38
39 TELECOMMUNICATION INFORMATION:
40
41 TELEPHONE: 202-628-5197
42
43 TELEFAX: 202-737-3528
44
45 INFORMATION FOR SEQ. ID NO: 1:
46
47 SEQUENCE CHARACTERISTICS:
48
49 LENGTH: 15 amino acids
50
51 TYPE: amino acid
52
53 TOPOLOGY: linear
54
55 MOLECULE TYPE: protein
56
57 US-08-732-371A-1

```

Query Match	100.0%	Score 62	DB 11	Length 15
Best Local Similarity	100.0%	Pred. No. 0.00053		
Matches 12	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	VEVPGSQHDSQ	12	
Db	1	VEVPGSQHDSQ	12	

```

RESULT      4
US-09-786-648-4
: Sequence 4, Application US/097866648
: GENERAL INFORMATION:
: APPLICANT: Williams, Neil Andrew
: APPLICANT: Hirst, Timothy Raymond
: TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vacci
: TITLE OF INVENTION: Adjuncts
: FILE REFERENCE: 7438
: CURRENT APPLICATION NUMBER: US/09/786,648
: CURRENT FILING DATE: 2001-03-07
: PRIOR APPLICATION NUMBER: PCT/GB99/02970
: PRIOR FILING DATE: 1999-09-07
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: MS DOS
: SEQ ID NO 4
: LENGTH: 21
: TYPE: PRT
: ORGANISM: E. coli
: FEATURE:
: LOCATION: 45...65
: OTHER INFORMATION: Isolated or synthetic EcxB beta4-alpha2 loop fragment derivab
: US-09-786-648-4

```

Query March	100.0%;	Score 62;	DB 21;	Length 21;
Best Local Similarity	100.0%;	Pred. No. 0.00079;		
Matches 12; Conservative	0;	Mismatches	0;	Indels
				Gaps 0;
OY 1 VEPGSGHDSQ 12				
db 6 VEPGSGHDSQ 17				

RESULT 5
US-09-786-648-5
; Sequence 5, Application US/09786648
; GENERAL INFORMATION:
; APPLICANT: Williams, Neil Andrew
; APPLICANT: Hirst, Timothy Raymond
; TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccine
; FILE REFERENCE: 7438
; CURRENT APPLICATION NUMBER: US/09/786,648
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: PCT/GB99/02970
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: MS DOS
; SEQ ID NO 5
; LENGTH: 21
; TYPE: PRP
; ORGANISM: E. coli
; FEATURE:
; LOCATION: 45...65
; OTHER INFORMATION: Isolated or synthetic ElxB beta4-alpha2 loop fragment derivable f
; OTHER INFORMATION: porcine E. coli
US-09-786-648-5

Query Match 100.0%; Score 62; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.00079;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSOHDSQ 12
|||||
DB 6 VEPGSOHDSQ 17

RESULT 6
US-08-110-947-10
; Sequence 10, Application US/08110947
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J
; APPLICANT: STEIN, Penelope E
; APPLICANT: COCKLE, Stephen A
; APPLICANT: COHEN, Raymond P
; APPLICANT: KLEIN, Michel H
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: Suite 1203 Crystal Plaza I, 2001 Jefferson
; STREET: Davis Hwy.
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,947
; FILING DATE: 24-AUG-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FALLON, Charles W
; REGISTRATION NUMBER: 28,946
; REFERENCE/DOCKET NUMBER: 1038-303 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; TELEX: 89-9456 LUKPAT

; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-110-947-10

Query Match 100.0%; Score 62; DB 5; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSOHDSQ 12
|||||
DB 40 VEPGSOHDSQ 51

RESULT 7
US-08-110-947A-26
; Sequence 26, Application US/08110947A
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J
; APPLICANT: STEIN, Penelope E
; APPLICANT: COCKLE, Stephen A
; APPLICANT: COHEN, Raymond P
; APPLICANT: KLEIN, Michel H
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: Suite 1203 Crystal Plaza I, 2001 Jefferson
; STREET: Davis Highway
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,947A
; FILING DATE: 24-AUG-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FALLON, Charles W
; REGISTRATION NUMBER: 28,946
; REFERENCE/DOCKET NUMBER: 1038-303 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; TELEX: 89-9456 LUKPAT
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-110-947A-26

Query Match 100.0%; Score 62; DB 5; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSOHDSQ 12
|||||
DB 40 VEPGSOHDSQ 51

RESULT 8

US-08-251-121-26
; Sequence 26, Application US/08251121
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J.
; APPLICANT: STEIN, Penelope E.
; APPLICANT: COCKLE, Stephen A.
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; APPLICANT: ARMSTRONG, Glen D.
; APPLICANT: HAZES, Bart
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Slim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/251,121
; FILING DATE: 31-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/110,947
; FILING DATE: 24-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-335
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-251-121-26
Query Match 100.0%; Score 62; DB 6; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VEVPGSQHDSQ 12
DB 40 VEVPGSQHDSQ 51
RESULT 9
US-08-393-334-2
; Sequence 2, Application US/08393334
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YACCOB, Reza K.
; APPLICANT: ZEALY, Gavin R.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Expression Of Gene Products From
; TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Slim & McBurney
; STREET: 330 University Avenue, Suite 701
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada

ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/393,334
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-417
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 416-595-1155
; TELEFAX: 416-595-1163
; TELEX: 065-24567 Simbas
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-393-334-2
Query Match 100.0%; Score 62; DB 7; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.005;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VEVPGSQHDSQ 12
DB 50 VEVPGSQHDSQ 61
RESULT 10
US-08-782-832-15
; Sequence 15, Application US/08782832
; GENERAL INFORMATION:
; APPLICANT: ARNTZEN, Charles J.
; APPLICANT: MASON, Hugh S.
; APPLICANT: HAQ, Tariq A.
; TITLE OF INVENTION: PRODUCTION OF AN ORALLY IMMUNOGENIC
; TITLE OF INVENTION: BACTERIAL PROTEIN IN TRANSGENIC PLANTS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pravel, Hewitt, Kimball & Krieger
; STREET: 1477 West Loop South, 10th Floor
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77027-9095
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/782,832
; FILING DATE: 13-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/328,716
; FILING DATE: 24-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Kettelberger, Denise M.
; REGISTRATION NUMBER: 33,924
; REFERENCE/DOCKET NUMBER: 36170/3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-850-0909
; TELEFAX: 713-850-0165
; INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-782-832-15

Query Match 100.0%; Score 62; DB 11; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.005;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12
|||||
DB 50 VEVPGSQHDSQ 61

RESULT 11
US-08-817-906-21

; Sequence 21, Application US/08817906
; GENERAL INFORMATION:
; APPLICANT: Charles J. Arntzen, Hugh S. Mason, John D. Clements,
; APPLICANT: and Tariq A. Haq
; TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski L.L.P.
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,906
; FILING DATE: 08/04/97
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13376
; FILING DATE: 24-OCT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fox, David L.
; REGISTRATION NUMBER: 40,612
; REFERENCE/DOCKET NUMBER: 36170/3P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-651-5151
; TELEFAX: 713-651-5246
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-817-906-21

Query Match 100.0%; Score 62; DB 12; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.005;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12
|||||
DB 50 VEVPGSQHDSQ 61

RESULT 12
US-09-836-433-14
; Sequence 14, Application US/09836433
; GENERAL INFORMATION:
; APPLICANT: Yuki, Yoshikazu
; APPLICANT: Uda, Shigezo

; TITLE OF INVENTION: HYBRID PROTEINS FOR AUTOIMMUNE DISEASE
; FILE REFERENCE: Not Assigned
; CURRENT APPLICATION NUMBER: US/09/836,433
; CURRENT FILING DATE: 2001-04-16
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 14
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Vibrio cholerae
US-09-836-433-14

Query Match 100.0%; Score 62; DB 22; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.005;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12
|||||
DB 50 VEVPGSQHDSQ 61

RESULT 13
US-09-836-433-20

; Sequence 20, Application US/09836433
; GENERAL INFORMATION:
; APPLICANT: Yuki, Yoshikazu
; APPLICANT: Uda, Shigezo
; TITLE OF INVENTION: HYBRID PROTEINS FOR AUTOIMMUNE DISEASE
; FILE REFERENCE: Not Assigned
; CURRENT APPLICATION NUMBER: US/09/836,433
; CURRENT FILING DATE: 2001-04-16
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 116
; TYPE: PRT
; ORGANISM: synthetic construct
US-09-836-433-20

Query Match 100.0%; Score 62; DB 22; Length 116;
Best Local Similarity 100.0%; Pred. No. 0.005;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12
|||||
DB 50 VEVPGSQHDSQ 61

RESULT 14
US-09-836-433-22

; Sequence 22, Application US/09836433
; GENERAL INFORMATION:
; APPLICANT: Yuki, Yoshikazu
; APPLICANT: Uda, Shigezo
; TITLE OF INVENTION: HYBRID PROTEINS FOR AUTOIMMUNE DISEASE
; FILE REFERENCE: Not Assigned
; CURRENT APPLICATION NUMBER: US/09/836,433
; CURRENT FILING DATE: 2001-04-16
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 119
; TYPE: PRT
; ORGANISM: synthetic construct
US-09-836-433-22

Query Match 100.0%; Score 62; DB 22; Length 119;
Best Local Similarity 100.0%; Pred. No. 0.005;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSQHDSQ 12
 |||||
 Db 50 VEVPGSQHDSQ 61

RESULT 15
 PCT-US99-30747-55
 ; Sequence 55, Application PC/TUS9930747
 ; GENERAL INFORMATION:
 ; APPLICANT: Boyce Thompson Institute for Plant Research at Cor
 ; TITLE OF INVENTION: Orally Immunogenic Bacterial Enterotoxins Expressed In
 ; TITLE OF INVENTION: Transgenic Plants
 ; FILE REFERENCE: 4868/85427
 ; CURRENT APPLICATION NUMBER: PCT/US99/30747
 ; CURRENT FILING DATE: 1999-12-22
 ; NUMBER OF SEQ ID NOS: 67
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 55
 ; LENGTH: 124
 ; TYPE: PRT
 ; ORGANISM: Escherichia coli
 PCT-US99-30747-55

Query Match 100.0%; Score 62; DB 1; Length 124;
 Best Local Similarity 100.0%; Pred. No. 0.0062;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 VEVPGSQHDSQ 12
 |||||
 Db 71 VEVPGSQHDSQ 82

Search completed: July 3, 2002, 09:03:20
 Job time: 620 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 09:04:12 ; Search time 46.52 Seconds
(without alignments)
26.099 Million cell updates/sec

Title: US-09-786-648-3

Perfect score: 62

Sequence: 1 VEPVGSQHDSQ 12

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5
308740 seqs, 101176262 residues

Total number of hits satisfying chosen parameters: 308740

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA.New*

1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	100.0	103	6	US-10-110-364-8
2	62	100.0	103	6	US-10-110-364-10
3	62	100.0	103	6	US-10-110-364-13
4	62	100.0	103	6	US-10-110-364-12
5	62	100.0	105	6	US-10-110-364-22
6	62	100.0	123	6	US-10-110-364-17
7	62	100.0	123	6	US-10-110-364-20
8	62	100.0	124	6	US-10-110-364-2
9	62	100.0	124	6	US-10-110-364-6
10	62	100.0	124	6	US-10-110-364-11
11	62	100.0	124	6	US-10-110-364-16
12	62	100.0	124	6	US-10-110-364-18
13	62	100.0	124	6	US-10-110-364-19
14	62	100.0	124	6	US-10-110-364-21
15	62	100.0	124	6	US-10-110-364-23
16	62	100.0	125	6	US-10-110-364-15
17	62	100.0	138	6	US-10-141-627-2
18	56	90.3	103	6	US-10-110-364-5
19	56	90.3	103	6	US-10-110-364-7
20	56	90.3	124	6	US-10-110-364-4
21	56	90.3	124	6	US-10-110-364-9
22	56	90.3	382	5	US-09-809-033A-3
23	56	90.3	382	5	US-09-809-033A-4
24	42	67.7	2359	6	US-10-155-881-28520
25	39	62.9	412	5	US-09-978-403A-157
26	39	62.9	412	5	US-09-978-544A-157

27	39	62.9	412	5	US-09-978-681A-157	Sequence 157, App
28	39	62.9	412	5	US-09-978-757A-157	Sequence 157, App
29	39	62.9	412	5	US-09-978-564A-157	Sequence 157, App
30	39	62.9	412	5	US-09-999-831A-157	Sequence 157, App
31	39	62.9	412	5	US-09-999-829A-157	Sequence 157, App
32	39	62.9	412	5	US-09-978-375A-157	Sequence 157, App
33	39	62.9	412	5	US-09-978-423A-157	Sequence 157, App
34	39	62.9	412	6	US-10-013-922A-157	Sequence 157, App
35	39	62.9	412	6	US-10-013-929A-157	Sequence 157, App
36	39	62.9	412	6	US-10-013-918A-157	Sequence 157, App
37	39	62.9	412	6	US-10-017-082A-157	Sequence 157, App
38	39	62.9	412	6	US-10-017-085A-157	Sequence 157, App
39	39	62.9	412	6	US-10-013-916A-157	Sequence 157, App
40	39	62.9	412	6	US-10-017-086A-157	Sequence 157, App
41	39	62.9	412	6	US-10-013-923A-157	Sequence 157, App
42	39	62.9	412	6	US-10-017-081A-157	Sequence 157, App
43	39	62.9	412	6	US-10-016-177A-157	Sequence 157, App
44	39	62.9	412	6	US-10-017-084A-157	Sequence 157, App
45	39	62.9	412	6	US-10-013-923A-157	Sequence 157, App

ALIGNMENTS

RESULT 1
US-10-110-364-8
; Sequence 8, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; PRIOR FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 8
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(103)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 998409
US-10-110-364-8

Query Match 100.0%; Score 62; DB 6; Length 103;
Best Local Similarity 100.0%; Pred No. 0.00021;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPVGSQHDSQ 12
Db 50 VEPVGSQHDSQ 61

RESULT 2
US-10-110-364-10
; Sequence 10, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364

```

; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(103)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 1421511
US-10-110-364-10

Query Match          100.0%; Score 62; DB 6; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.00021;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSHIDSO 12
   |||||
Db 50 VEPGSHIDSO 61

RESULT 3
US-10-110-364-13
; Sequence 13, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: AB5 TOXIN B SUBUNIT MUTANTS WITH ALTERED
; TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(103)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 14215235
US-10-110-364-13

Query Match          100.0%; Score 62; DB 6; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.00021;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSHIDSO 12
   |||||
Db 50 VEPGSHIDSO 61

RESULT 4
US-10-110-364-22
; Sequence 22, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
```

```

; TITLE OF INVENTION: AB5 TOXIN B SUBUNIT MUTANTS WITH ALTERED
; TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; PRIOR FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(103)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 494265.
US-10-110-364-22

Query Match          100.0%; Score 62; DB 6; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.00021;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSHIDSO 12
   |||||
Db 50 VEPGSHIDSO 61

RESULT 5
US-10-110-364-12
; Sequence 12, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: AB5 TOXIN B SUBUNIT MUTANTS WITH ALTERED
; TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(105)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 2781121 (Ogawa
US-10-110-364-12

Query Match          100.0%; Score 62; DB 6; Length 105;
Best Local Similarity 100.0%; Pred. No. 0.00021;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSHIDSO 12
   |||||
Db 51 VEPGSHIDSO 62

RESULT 6
US-10-110-364-17
; Sequence 17, Application US/10110364
```

```

; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(123)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 1395122.
US-10-110-364-17

```

```

Query Match      100.0%; Score 62; DB 6; Length 123;
Best Local Similarity 100.0%; Pred. No. 0.00025;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 VEPGSOHDSQ 12
    |||||
DB 71 VEPGSOHDSQ 82

```

```

RESULT 7
US-10-110-364-20
; Sequence 20, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(123)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 223254.
US-10-110-364-20

```

```

Query Match      100.0%; Score 62; DB 6; Length 123;
Best Local Similarity 100.0%; Pred. No. 0.00025;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VEPGSOHDSQ 12
    |||||
DB 70 VEPGSOHDSQ 81

```

```

RESULT 8
US-10-110-364-2
; Sequence 2, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Vibrio cholera
US-10-110-364-2

```

```

Query Match      100.0%; Score 62; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00026;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 VEPGSOHDSQ 12
    |||||
DB 71 VEPGSOHDSQ 82

```

```

RESULT 9
US-10-110-364-6
; Sequence 6, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 808900.
US-10-110-364-6

```

```

Query Match      100.0%; Score 62; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00026;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VEPGSOHDSQ 12
    |||||
DB 71 VEPGSOHDSQ 82

```

RESULT 10

US-10-110-364-11
; Sequence 11, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 48890 (classic)
; OTHER INFORMATION: 569B).
US-10-110-364-11

Query Match 100.0%; Score 62; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00026;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHIDSQ 12
|||||
DB 71 VEVPGSQHIDSQ 82

RESULT 11
US-10-110-364-16
; Sequence 16, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 1169505.
US-10-110-364-16

Query Match 100.0%; Score 62; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00026;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHIDSQ 12
|||||

DB 71 VEVPGSQHIDSQ 82

RESULT 12
US-10-110-364-18
; Sequence 18, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 145833.
US-10-110-364-18

Query Match 100.0%; Score 62; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00026;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHIDSQ 12
|||||
DB 71 VEVPGSQHIDSQ 82

RESULT 13
US-10-110-364-19
; Sequence 19, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 1648865
US-10-110-364-19

Query Match 100.0%; Score 62; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00026;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VEPGSHIDSQ 12
Db 71 VEPGSHIDSQ 82

RESULT 14

US-10-110-364-21
; Sequence 21, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: AB5 TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 408996.
US-10-110-364-21

Query Match 100.0%; Score 62; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00026;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSHIDSQ 12
Db 71 VEPGSHIDSQ 82

RESULT 15
US-10-110-364-23
; Sequence 23, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: AB5 TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 69630.
US-10-110-364-23

Query Match 100.0%; Score 62; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00026;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VEPGSHIDSQ 12
Db 71 VEPGSHIDSQ 82

Search completed: July 3, 2002, 09:04:12
Job time: 652 sec

Mon Jul 8 07:51:49 2002

us-09-786-648-3.rapn

Page 6

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 08:56:32 ; Search time 46.57 Seconds

(without alignments)
24.760 Million cell updates/sec

Title: US-09-786-648-3

Sequence: 1 VEPYGSQHDSQ 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.71.*
2: PIR1.*
3: PIR2.*
4: PIR3.*
5: PIR4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	62	100.0	124	1	XVVCB	Cholera enterotoxin
2	62	100.0	124	1	QLECB	heat-labile entero
3	40	64.5	242	1	T34767	hypothetical prote
4	39	62.9	260	2	C96837	protein P20B17.2 f
5	39	62.9	272	2	T14755	hypothetical prote
6	39	62.9	353	2	H75446	(S)-2-hydroxy-acid
7	38	61.3	282	1	A35781	hIppocampus-derive
8	38	61.3	674	1	A41670	carbon-monoxide de
9	37	59.7	352	2	H85518	hypothetical prote
10	37	59.7	367	2	T01751	glibberellin 20-oxi
11	37	59.7	376	1	S17246	chorismate synthas
12	37	59.7	378	2	F90667	hypothetical prote
13	37	59.7	423	2	G96554	hypothetical prote
14	37	59.7	442	2	T39683	zuooin-like protei
15	37	59.7	550	2	T01770	hypothetical prote
16	36	58.1	368	2	F98282	histidinol-phospha
17	36	58.1	368	2	A83001	histidinol-phospha
18	36	58.1	373	2	T47115	probable 4-carboxy
19	36	58.1	455	1	A69753	glucarate dehydrat
20	36	58.1	755	2	D95842	probable beta-gala
21	35	56.5	1785	2	S53976	probable membrane
22	35	56.5	142	2	S28757	cytochrome-c oxida
23	35	56.5	259	2	B84125	hydrolyase BH805 f
24	35	56.5	284	2	F82031	methylenetetrahydr
25	35	56.5	302	2	JN0143	protein B0495.8 f1
26	35	56.5	313	2	E88216	hypothetical prote
27	35	56.5	360	2	A90029	hypothetical prote
28	35	56.5	367	2	G83368	hypothetical prote
29	35	56.5	371	2	S23861	hypothetical prote

30	35	56.5	377	2	T29750	hypothetical prote
31	35	56.5	389	2	A86303	hypothetical prote
32	35	56.5	389	2	A82980	monooxygenase fimp
33	35	56.5	427	2	JC5694	stress-activated p
34	35	56.5	443	2	T29029	hypothetical prote
35	35	56.5	472	2	AG0432	glutamate synthase
36	35	56.5	539	2	S53529	monophenol monooxy
37	35	56.5	593	2	T24379	hypothetical prote
38	35	56.5	670	2	T02092	beta-fructofuranos
39	35	56.5	960	2	A41638	chitin synthase (E
40	35	56.5	961	2	A55380	faciogenital dyspl
41	35	56.5	1258	2	T29041	hypothetical prote
42	35	56.5	2206	2	G71611	hypothetical prote
43	34	54.8	168	2	T27563	hypothetical prote
44	34	54.8	181	2	T45990	hypothetical prote
45	34	54.8	244	1	T01091	NADH dehydrogenase

ALIGNMENTS

```
RESULT 1
XVVCB
cholera enterotoxin, B chain precursor VC1456 [validated] - Vibrio cholerae (strain N
M:Alternates names: enterotoxin beta chain
C:Species: Vibrio cholerae
C:Date: 24-Apr-1984 #sequence-revision 01-Sep-2000 #text-change 02-Feb-2001
C:Accession: S14624; S39238; S39241; H82196; JCI078; S17666; PC1010; A05130; A01819;
R:Dams, E.; de Wolf, M.; Dierick, W.
submitted to the EMBL Data Library, March 1991
A:Description: Correction of the cholera toxin nucleotide sequence of the Vibrio chol
A:Reference number: S14624
A:Accession: S14624
A:Molecule type: DNA
A:Residues: 1-124 <DAM>
A:Cross-references: EMBL:X58786; NID:948420; PIDN:CAA41593.1; PID:948422
A:Experimental source: Strain 2125
R:Lebens, M.; Holmgren, J.
submitted to the EMBL Data Library, November 1993
A:Description: Structure and arrangement of the Cholera toxin genes in Vibrio Cholera
A:Reference number: S39238
A:Accession: S39238
A:Molecule type: DNA
A:Residues: 1-124 <DAM>
A:Cross-references: EMBL:X76390; NID:9433856; PIDN:CAA53973.1; PID:9433857
A:Residues: 1-124 <LEB>
A:Cross-references: EMBL:X76390; NID:9433856; PIDN:CAA53973.1; PID:9433857
A:Accession: S39241
A:Molecule type: DNA
A:Residues: 1-124 <LEW>
A:Cross-references: EMBL:X76391; NID:9433859; PIDN:CAA53976.1; PID:9433861
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
Chadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bess, S.; Qin, H.; Dragoi, I.; Sellers
L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: H82196
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-124 <HEI>
A:Cross-references: GB:AE004224; GB:AE003852; NID:99655952; PIDN:AAF4613.1; GSPDB:GN
A:Experimental source: serogroup O1, strain N16961; biotype El Tor
R:Shi, C.H.; Cao, C.; Zhang, J.S.; Ma, Q.J.
Chinese Biochem. J. 9, 395-399, 1993
A:Title: Nucleotide sequence analysis of the gene encoding the classical biotype V.ch
A:Reference number: JCI078
A:Accession: JCI078
A:Molecule type: DNA
A:Residues: 1-20, 'Q', 22-31, 'Q', 33-38, 'H', 40-49, 'G', 51-67, 'T', 69-124 <SHI>
A:Experimental source: classical biotype strain 569b
R:Dams, E.; de Wolf, M.; Dierick, W.
Biochim. Biophys. Acta 1090, 139-141, 1991
A:Title: Nucleotide sequence analysis of the CT operon of the Vibrio cholerae class
A:Reference number: S17665; MUID:91355224
```

A:Accession: S17666
A:Molecule type: DNA
A:Residues: 1-38, 'H', 40-67, 'T', 69-124 <DA2>
A:Cross-references: EMBL:X58785; NID:948888; PIDN:CAA41591.1; PID:948890
R:Ma, Q.J.; Liu, C.X.; Xiong, L.S.; Yu, X.Q.
Sci. Sin. B Chem. Biol. Agric. Med. Earth Sci. 34, 274-280, 1991
A:Title: B subunit of cholera toxin produced in *Escherichia coli*.
A:Reference number: PC1010
A:Accession: PC1010
A:Molecule type: protein
A:Residues: 22-38, 'H', 40-41 <MAO>
R:Meekalanos, J.J.; Swartz, D.J.; Pearson, G.D.N.; Harford, N.; Groyne, F.; de Wilde, M.
Nature 306, 551-557, 1983
A:Reference number: A93320; MWID:84068199
A:Accession: A05130
A:Molecule type: DNA
A:Residues: 1-32, 'S', 34-74, 'S', 76-124 <MEK>
A:Cross-references: GB:X00171; NID:948347; PIDN:CAA2496.1; PID:g758351
R:Kurosky, A.; Markel, D.E.; Peterson, J.W.
J. Biol. Chem. 252, 7257-7264, 1977
A:Title: Covalent structure of the beta chain of cholera enterotoxin.
A:Reference number: A01819; MWID:78005537
A:Accession: A01819
A:Molecule type: protein
A:Residues: 22-38, 'H', 40-42, 'N', 44-67, 'T', 69-90, 'N', 92-124 <KUR>
R:Uai, C.Y.
J. Biol. Chem. 252, 7249-7256, 1977
A:Title: Determination of the primary structure of cholera toxin B subunit.
A:Reference number: A38033; MWID:78005536
A:Accession: A38033
A:Molecule type: protein
A:Residues: 22-38, 'H', 40-42, 'N', 44-67, 'T', 69, 'E', 71-90, 'N', 92-124 <LA1>
A:Note: The difference at residue 70 may be due to deamidation during preparation
R:Nakshima, Y.; Napietkowski, P.; Schaffer, D.E.; Konigsberg, W.H.
FEBS Lett. 68, 275-278, 1976
A:Title: Primary structure of the B subunit of cholera enterotoxin.
A:Reference number: A38034; MWID:77026365
A:Accession: A38034
A:Molecule type: protein
A:Residues: 22-38, 'H', 40-67, 'T', 69, 'E', 71, 'Q', 74-75, 'VE', 78-86, 'Q', 88-99, 'Q', 101-103, 'Q'
R:Takao, T.; Watanabe, H.; Shimonishi, Y.
Eur. J. Biochem. 146, 503-508, 1985
A:Title: Facile identification of protein sequences by mass spectrometry.
A:Reference number: A21910; MWID:85126976
A:Accession: A21910
A:Molecule type: protein
A:Residues: 22-38, 'H', 40-42, 'N', 44-67, 'T', 69-90, 'N', 92-124 <TAK>
A:Experimental source: biotype Inaba 569B
A:Note: Asn-65 was partially deaminated to Asp
C:Comment: The authors translated the codon TCA for residue 33 as Tyr.
C:Genetics:
A:Gene: VC1456
A:Map position: 1
C:Complex: the cholera enterotoxin molecule contains three kinds of chains; an alpha and
beta noncovalently with the subunit B, an aggregate of five beta chains
C:Function:
A:Description: involved in binding of the toxin to cell membranes
C:Superfamily: cholera enterotoxin beta chain
C:Keywords: enterotoxin; toxin
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-124/Product: cholera enterotoxin chain B #status predicted <MAT>
F:30-107/Disulfide bonds: #status experimental

Query Match 100.0%; Score 62; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00019;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSOHIDSO 12
|||||
DB 71 VEVPGSOHIDSO 82

RESULT 2

OLECB
heat-labile enterotoxin chain B precursor - *Escherichia coli*
C:Species: *Escherichia coli*
C:Date: 29-Jun-1981 #sequence, revision 29-Jun-1981 #text, change 18-Jun-1999
C:Accession: A01820; B26946; I41194; I41287; I67644; A61475
R:Dallas, W.S.; Falkow, S.
Nature 288, 499-501, 1980
A:Title: Amino acid sequence homology between cholera toxin and *Escherichia coli* heat
A:Reference number: A01820; MWID:81074965
A:Accession: A01820
A:Molecule type: mRNA
A:Residues: 1-124 <DAL>
R:Yamamoto, T.; Gojobori, T.; Yokota, T.
J. Bacteriol. 169, 1352-1357, 1987
A:Title: Evolutionary origin of pathogenic determinants in enterotoxigenic *Escherichia*
A:Reference number: A26946; MWID:87137303
A:Accession: B26946
A:Molecule type: DNA
A:Residues: 1-27, 'E', 29-63, 'K', 65-124 <YAM>
A:Cross-references: EMBL:M17874; NID:g145830; PIDN:AAA24792.1; PID:g148336
R:Leong, J.; Vinal, A.C.; Dallas, W.S.
Infect. Immun. 48, 73-77, 1985
A:Title: Nucleotide sequence comparison between heat-labile toxin B-subunit cistrons
A:Reference number: I41194; MWID:85156481
A:Accession: I41194
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5, 'F', 7-17, 'C', 19-24, 'S', 26-27, 'E', 29-33, 'H', 35-63, 'K', 65-66, 'A', 68-122
A:Cross-references: GB:M17874; NID:g145830; PIDN:AAA24792.1; PID:g145831
R:Ibrahim, I.; Gentz, R.
J. Biol. Chem. 262, 10189-10194, 1987
A:Title: A functional interaction between the signal peptide and the translation appa
ticulum.
A:Reference number: I41287; MWID:87280041
A:Accession: I41287
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-22 <RE2>
A:Cross-references: GB:M17101; NID:g146375; PIDN:AAA23973.1; PID:g146376
R:Inoue, T.; Tsuji, T.; Koto, M.; Imamura, S.; Miyama, A.
FEBS Microbiol. Lett. 108, 157-161, 1993
A:Title: Amino acid sequence of heat-labile enterotoxin from chicken enterotoxigenic
A:Reference number: I53542; MWID:93252225
A:Accession: I57644
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-17, 'C', 19, 'Y', 21-24, 'S', 26-27, 'E', 29-63, 'K', 65-66, 'A', 68-122, 'E', 124 <R
A:Cross-references: GB:S60731; NID:g408994; PIDN:AAC60441.1; PID:g408996
R:Tsuji, T.; Lida, T.; Honda, T.; Mawatani, T.; Negahama, M.; Sakurai, J.; Wada, K.;
Microb. Pathog. 2, 381-390, 1987
A:Title: A unique amino acid sequence of the B subunit of a heat-labile enterotoxin I
A:Reference number: A61475; MWID:89180953
A:Accession: A61475
A:Molecule type: protein
A:Residues: 22-24, 'S', 26-27, 'E', 29-63, 'K', 65-66, 'A', 68-95, 'A', 97-122, 'E', 124 <TSU>
A:Experimental source: strain 240-3
C:Complex: the heat-labile enterotoxin molecule contains one A chain and five or six
C:Function:
A:Description: the biological activity of the toxin is produced by the A chain, which
C:Superfamily: cholera enterotoxin beta chain
C:Keywords: enterotoxin
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-124/Product: heat-labile enterotoxin chain B #status predicted <MAT>
F:30-107/Disulfide bonds: #status predicted

Query Match 100.0%; Score 62; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00019;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSOHIDSO 12

Db 71 VEVPGSOHIDSQ 82

RESULT 3

T34767 hypothetical protein SC2A11.21c SC2A11.21c - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999

C:Accession: T34767

R:Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

A:Reference number: 221556

A:Accession: T34767

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-242 <MUR>

A:Cross-references: EMBL:AL031184; PIDN:CAA20190.1; GSPDB:GN00070; SCOEDB:SC2A11.21c

C:Genetics:

A:gene: SCOEDB:SC2A11.21c

Query Match 64.5%; Score 40; DB 2; Length 242;
Best Local Similarity 70.0%; Pred. No. 6.1;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 VEVPGSOHID 10

Db 204 VEVPGTHDFTD 213

RESULT 4

C96827 protein P20B17.2 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: C96827

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.M.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

anssen, N.F.; Hughes, B.; Hultzer, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marzalli,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: C96827

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-260 <STO>

A:Cross-references: GB:AE005173; NID:g7715588; PIDN:AAF6106.1; GSPDB:GN00141

C:Genetics:

A:gene: P20B17.2

A:Map position: 1

Query Match 62.9%; Score 39; DB 2; Length 260;
Best Local Similarity 50.0%; Pred. No. 10;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 VEVPGSOHIDSQ 12

Db 65 VEVPGKOHVSEK 76

RESULT 5
T14755 hypothetical protein DKFZp564A0122.1 - human
C:Species: Homo sapiens (man)
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T14755
R:Mambuli, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, August 1999

A:Reference number: Z18181

A:Accession: T14755

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-272 <MAN>

A:Cross-references: EMBL:AL110209

A:Experimental source: fetal brain; clone DKFZp564A0122

C:Genetics:

A:Note: DKFZp564A0122.1

Query Match 62.9%; Score 39; DB 2; Length 272;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 EVPGSOHID 10

Db 246 ELPGESEHIE 254

RESULT 6

H75446 (S)-2-hydroxy-acid oxidase - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000

C:Accession: H75446

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.

; M.; Shen, M.; Yamathavan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.;

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896

A:Accession: H75446

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-353 <MHI>

A:Cross-references: GB:AE001954; GB:AE000513; NID:g6458751; PIDN:AAF10604.1; PID:g645

A:Experimental source: strain R1

C:Genetics:

A:gene: DR1031

A:Map position: 1

C:Superfamily: (S)-2-hydroxy-acid oxidase; (S)-2-hydroxy-acid oxidase homology

F;3-297/Domain: (S)-2-hydroxy-acid oxidase homology <2HY>

Query Match 62.9%; Score 39; DB 2; Length 353;
Best Local Similarity 75.0%; Pred. No. 14;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 3 VPGSOHID 10

Db 189 VPGSEHID 196

RESULT 7
A35781 hippocampus-derived neurotrophic factor precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 05-Oct-1990 #sequence_revision 05-Oct-1990 #text_change 16-Jul-1999

C:Accession: A35781; A40094

R:Enfors, P.; Ibanez, C.F.; Ebendal, T.; Olson, L.; Persson, H.

Proc. Natl. Acad. Sci. U.S.A. 87, 5454-5458, 1990

A:Title: Molecular cloning and neurotrophic activities of a protein with structural

A:Reference number: A35781; MUID:90319130

A:Accession: A35781

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-282 <ERN>

A:Cross-references: GB:M34643

R;Maisonnier, P.C.; Belluscio, L.; Squinto, S.; Ip, N.Y.; Firth, M.E.; Lindsay, R.M.; Science 247, 1446-1451, 1990
A:Title: Neurotrophin-3: a neurotrophic factor related to NGF and BDNF.
A:Reference number: A40094; MUID:90208301
A:Accession: A40094
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 25-282 <MAI>
A:Cross-references: GB:M33968; NID:g205771; PIDN:AAA1727.1; PID:g205772
C:Superfamily: nerve growth factor beta chain

Query Match 61.3%; Score 38; DB 2; Length 282;
Best Local Similarity 54.5%; Pred. No. 17;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
OY 1 VEPGSHIDS 11
Db 1 VDPGNSHTDA 11

RESULT 8
A41670
Carbon-monoxide dehydrogenase (EC 1.2.99.2) beta chain [similarity] - Clostridium thermac
C:Species: Clostridium thermacellum
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000
C:Accession: A41670
R:Morton, T.A.; Rungquist, J.A.; Ragsdale, S.W.; Shanmugasundaram, T.; Wood, H.G.; Ljung
J. Biol. Chem. 266, 23824-23828, 1991
A:Title: The primary structure of the subunits of carbon monoxide dehydrogenase/acetyl-C
A:Reference number: A41670; MUID:92084676
A:Accession: A41670
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-674 <MOR>
A:Cross-references: GB:M62727; NID:g144784; PIDN:AAA23228.1; PID:g144785
C:Superfamily: carbon-monoxide dehydrogenase beta chain; hybrid cluster [4Fe-2S-30] homoc
C:Keywords: 4Fe-2S-30; 4Fe-4S; electron transfer; iron-sulfur protein; metalloprotein; c
F:256-606/Domain: hybrid cluster [4Fe-2S-30] homology <HCL>
F:68,71,76,90/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted
F:283,321,355,470,500,550,585/Binding site: Ni-3Fe-2S-30 cluster (His, Glu, Cys, Cys, Cy
F:470/Modified site: cysteine persulfide (Cys) #status predicted

Query Match 61.3%; Score 38; DB 1; Length 674;
Best Local Similarity 54.5%; Pred. No. 45;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 EVPGSHIDSQ 12
Db 379 KPGAYHIDYQ 389

RESULT 9
B85518
hypothetical protein Z0347 [imported] - Escherichia coli (strain O157:H7, substrain EDLg
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: B85518
R:Penna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Hiller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoukis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85518
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-352 <STO>
A:Cross-references: GB:AE005174; NID:g12513064; PIDN:AAG54606.1; GSPDB:GN00145; UWGP:Z03
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z0347

Query Match 59.7%; Score 37; DB 2; Length 352;
Best Local Similarity 60.0%; Pred. No. 35;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 VEPGSHID 10
Db 320 VVPGANHVD 329

RESULT 10
T01751
gibberellin 20-oxidase - common tobacco
N:Alternate names: Ntcl6 protein
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 20-Jun-2000
C:Accession: T01751
R:Tanaka-Ueguchi, M.; Itch, H.; Oyama, N.; Koshioka, M.; Matsuoka, M.
submitted to the EMBL Data Library, July 1998
A:Description: Over-expression of a tobacco homeobox gene, NTH15, decreases the expe
A:Reference number: Z14418
A:Accession: T01751
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-367 <TAN>
A:Cross-references: EMBL:AB016084
C:Genetics:
A:Gene: Ntcl6
C:Superfamily: 1-aminocyclopropane-1-carboxylate oxidase

Query Match 59.7%; Score 37; DB 2; Length 367;
Best Local Similarity 87.5%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 EVPGSHI 9
Db 149 EVPGSHI 156

RESULT 11
S17246
choistmate synthase (EC 4.6.1.4) - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein G2501; protein YGL148w
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: S17246; S64162
R:Jones, D.G.L.; Reusser, U.; Baus, G.H.
Mol. Microbiol. 5, 2143-2152, 1991
A:Title: Molecular cloning, characterization and analysis of the regulation of the AR
A:Reference number: S17246; MUID:92114793
A:Accession: S17246
A:Molecule type: DNA
A:Residues: 1-376 <JON>
A:Cross-references: EMBL:X60190; NID:g3386; PIDN:CAA42745.1; PID:g3387
R:Volckert, G.; Voel, M.; Vernasselt, P.; Defoor, E.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64153
A:Accession: S64162
A:Molecule type: DNA
A:Residues: 1-376 <VOL>
A:Cross-references: EMBL:Z72670; NID:g1322731; PIDN:CAA96860.1; PID:g1322732; GSPDB:G
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:ARO2; MIPS:YGL148w
A:Cross-references: SGD:S0003116; MIPS:YGL148w
A:Map position: 7L
C:Superfamily: choistmate synthase
C:Keywords: phosphorus-oxygen lyase; transmembrane protein
F:347-363/Domain: transmembrane #status predicted <TMW>

Query Match 59.7%; Score 37; DB 1; Length 376;

Best Local Similarity 70.0%; Pred. No. 37;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 VEPGSOHID 10
Db 264 VVPGANHVD 273

RESULT 12

hypothetical protein ECS0310 [imported] - Escherichia coli (strain O157:H7, substrain R1)
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: F90667
R:Rayashli, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gatawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MID:21156231; PMID:11258796
A:Accession: F90667
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-378 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA033733.1; PID:q13359767; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: ECS0310

Query Match 59.7%; Score 37; DB 2; Length 378;
Best Local Similarity 60.0%; Pred. No. 37;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 VEPGSOHID 10
Db 346 VVPGANHVD 355

RESULT 13

G96554
hypothetical protein F19C24.16 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G96554
R:Theologos, A.; Eckert, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Mailli, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MID:21016719
A:Accession: G96554
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-423 <STO>
A:Cross-references: GB:AE005173; MID:g11094753; PIDN:AGC29686.1; GSPDB:GN00141
C:Genetics:
A:Gene: F19C24.16
A:Map position: 1

Query Match 59.7%; Score 37; DB 2; Length 423;
Best Local Similarity 75.0%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 VPGSOHID 10
Db 403 VPGIOHVD 410

RESULT 14

zucotin-like protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C:Accession: T39683
R:Oliver, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, March 1998
A:Reference number: Z21869
A:Accession: T39683
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-124 <OLI>
A:Cross-references: EMBL:AL049489; PIDN:CAB39796.1; GSPDB:GN00067; SPDB:SPBC1778.01c
A:Experimental source: strain 972h-; cosmid c1778
R:Mod, V.; Rajandream, M.A.; Barrell, B.G.; Lauber, J.; Hilbert, H.; Duesterhoeft, A.
submitted to the EMBL Data Library, February 1998
A:Reference number: Z21910
A:Accession: T40195
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 89-442 <WOO>
A:Cross-references: EMBL:Z97992; PIDN:CAB10796.1; GSPDB:GN00067; SPDB:SPBC30D10.01
C:Genetics:
A:Gene: SPDB:SPBC1778.01c; SPDB:SPBC30D10.01
A:Map position: 2

Query Match 59.7%; Score 37; DB 2; Length 442;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 2 EVPGSOHIDS 11
Db 371 DVPSAHEHVS 380

RESULT 15

T01770
hypothetical protein A_ICG002P16.9 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 22-Oct-1999
C:Accession: T01770
R:Miller, N.; Beck, C.; Kramer, J.
submitted to the EMBL Data Library, June 1997
A:Description: The sequence of A. thaliana ICG002P16.
A:Reference number: Z14421
A:Accession: T01770
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-550 <ML>
A:Cross-references: EMBL:AF007270; MID:g2191157; PID:g2191172; GSPDB:GN00063; ATSP:A.
C:Genetics:
A:Gene: ATSP:A_ICG002P16.9
A:Map position: 5
A:Introns: 159/1; 272/1; 434/2; 477/3

Query Match 59.7%; Score 37; DB 2; Length 550;
Best Local Similarity 66.7%; Pred. No. 56;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 VEPGSOHIDSQ 12
Db 260 VTCSGSOHIDFQ 271

Search completed: July 3, 2002, 08:56:32
Job time: 232 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 09:04:41 ; Search time 21.51 Seconds

(without alignments)
21.601 Million cell updates/sec

Title: US-09-786-648-3
Perfect score: 62
Sequence: 1 VEPGSHIDSQ 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62	100.0	124	1 CHTB_VIBCH	P01556 vibrio chol
2	62	100.0	124	1 EUBH_ECOLI	P13811 escherichia
3	62	100.0	124	1 EUBP_ECOLI	P32890 escherichia
4	38	61.3	674	1 DCMB_MOOTH	P27989 moorella th
5	37	59.7	376	1 AROC_YEAST	P28777 saccharomyc
6	36	58.1	240	1 YB5A_THEMA	P58009 thermotoga
7	36	58.1	455	1 GUDH_BACSU	P42238 bacillus su
8	36	58.1	743	1 BGAL_THERT	P77989 thermoaer
9	36	58.1	855	1 SRI4_HUMAN	Q09596 homo sapien
10	36	58.1	1785	1 GLS3_YEAST	Q04952 saccharomyc
11	35	56.5	302	1 PHEB_PSESP	P31019 pseudomonas
12	35	56.5	313	1 YP68_CAEBL	O09217 caenorhabdi
13	35	56.5	371	1 YM18_PSEAE	O01609 pseudomonas
14	35	56.5	539	1 TYRO_ASPOR	O00234 aspergillus
15	35	56.5	670	1 INVI_MAIZE	P49175 zea mays (m
16	35	56.5	960	1 CHS3_NEUCR	P28070 neurospora
17	35	56.5	960	1 FGD1_MOUSE	P52734 mus musculu
18	35	56.5	961	1 FGD1_HUMAN	P98174 homo sapien
19	34	54.8	219	1 CIDA_HUMAN	O60543 homo sapien
20	34	54.8	244	1 NUHM_ARATH	O22769 arabidopsis
21	34	54.8	365	1 FTZ3_PYRAB	O94005 pyrococcus
22	34	54.8	365	1 FTZ3_PYRHO	O50060 pyrococcus
23	34	54.8	419	1 GSC_DROME	P54366 drosophila
24	34	54.8	432	1 AROC_NEUCR	O12640 neurospora
25	34	54.8	456	1 SR54_THEAC	O9hk10 thermoplasm
26	34	54.8	459	1 IL7R_MOUSE	P16872 mus musculu
27	34	54.8	491	1 CD5_RAT	P51882 rattus norv
28	34	54.8	500	1 GABT_HUMAN	P68044 homo sapien
29	34	54.8	508	1 Y202_HUMAN	Q92599 homo sapien
30	34	54.8	557	1 HLYB_SERMA	P15321 serattia ma
31	34	54.8	560	1 INRI_SHEEP	Q28589 ovis aries
32	34	54.8	775	1 THIL_SCHPO	P36598 schistosach
33	34	54.8	842	1 DPOL_HPBVM	P31870 hepatitis b

34	34	54.8	1151	1 Y245_TREPA	O83273 treponema p
35	33.5	54.0	375	1 MASP_MOUSE	P70124 mus musculu
36	33.5	54.0	375	1 MASP_RAT	P70564 rattus norv
37	33	53.2	259	1 MOB2_YEAST	P43563 saccharomyc
38	33	53.2	264	1 FTSQ_STRCU	O86038 streptomyc
39	33	53.2	291	1 CORC_BUCAI	P57518 buchiera ap
40	33	53.2	386	1 YVAN_BACSU	P37535 bacillus su
41	33	53.2	454	1 VNUC_THOCV	P89216 thogoto vir
42	33	53.2	504	1 A37C_DROME	P18487 drosophila
43	33	53.2	534	1 YOG1_CAEBL	P34610 caenorhabdi
44	33	53.2	621	1 HEMI_AGABI	O92403 agaricus bi
45	33	53.2	666	1 PD14_MOUSE	O92183 mus musculu

ALIGNMENTS

RESULT	ID	CHTB_VIBCH	STANDARD	PRT	124 AA.
AC	P01556	Q9J002			
DT	21-JUL-1986	(Rel. 01, Created)			
DT	13-AUG-1987	(Rel. 05, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Cholera enterotoxin, beta chain precursor.				
GN	CTXB OR TOXB OR VC1456.				
OS	Vibrio cholerae.				
OC	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.				
OX	NCBI_TaxID=666;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=84061784; PubMed=6315707;				
RA	Lockman H., Kaper J.B.;				
RT	"Nucleotide sequence analysis of the A2 and B subunits of vibrio				
RT	cholerae enterotoxin.";				
RL	J. Biol. Chem. 258:13722-13726(1983).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	STRAIN=EL TOR 2125;				
RX	MEDLINE=84068199; PubMed=6646234;				
RA	Mekalanos J.J., Swartz D.J., Pearson G.D.N., Harford N., Groyne F.,				
RA	de Wilde M.;				
RT	"Cholera toxin genes: nucleotide sequence, deletion analysis and				
RT	vaccine development.";				
RL	Nature 306:551-557(1983).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RX	STRAIN=EL TOR 2125;				
RA	Dans E., de Wolf M., Dierick W.;				
RA	Submitted (MAY-1991) to the EMBL/genbank/DBJ databases.				
RL	[4]				
RP	SEQUENCE FROM N.A.				
RX	STRAIN=4260B / SEROTYPE O139;				
RX	MEDLINE=94237453; PubMed=8181723;				
RA	Lebens M., Holmgren J.;				
RT	"Structure and arrangement of the cholera toxin genes in Vibrio				
RT	cholerae O139.";				
RL	FEMS Microbiol. Lett. 117:197-202(1994).				
RN	[5]				
RP	SEQUENCE FROM N.A.				
RX	STRAIN=1854 / O139-BENGAL;				
RA	Yamamoto K., Do V.G.R.F., Xu M., Iida T., Miwatani T., Albert M.J.,				
RA	Honda T.;				
RL	Submitted (MAY-1994) to the EMBL/genbank/DBJ databases.				
RN	[6]				
RP	SEQUENCE FROM N.A.				
RX	STRAIN=EL TOR N16961 / SEROTYPE O1;				
RX	MEDLINE=20406833; PubMed=10952301;				
RA	Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,				
RA	Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,				
RA	Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,				
RA	Ernolleva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,				
RA	McDonald L., Uitterback T., Fleischmann R.D., Niernan W.C., White O.,				

RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
 RT *cholerae*.";
 RL Nature 406:477-483(2000).
 RN [7]
 RP SEQUENCE OF 22-124.
 RX MEDLINE=78005357; PubMed=903363;
 RA Kurosky A., Markel D.E., Peterson J.W.;
 RT "Covalent structure of the beta chain of cholera enterotoxin.";
 RL J. Biol. Chem. 252:7257-7264(1977).
 RN [8]
 RP SEQUENCE OF 22-124.
 RX MEDLINE=78005356; PubMed=903362;
 RA Lai C.-Y.;
 RT "Determination of the primary structure of cholera toxin B subunit.";
 RL J. Biol. Chem. 252:7249-7256(1977).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RX MEDLINE=94272319; PubMed=8003954;
 RA Merritt E.A., Sarfaty S., van den Akker F., L'Hoir C., Martiat J.A.,
 RA Hol W.G.J.;
 RT "Crystal structure of cholera toxin B-pentamer bound to receptor GMI
 RT pentasaccharide.";
 RL Protein Sci. 3:166-175(1994).
 RN [10]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
 RX MEDLINE=95387394; PubMed=7658472;
 RA Zhang R.G., Westbrook M.L., Westbrook E.M., Scott D.L., Orlowski Z.,
 RA Maulik P.R., Reed R.A., Shipley G.G.;
 RT "The 2.4 A crystal structure of cholera toxin B subunit pentamer:
 RT choleragenoid.";
 RL J. Mol. Biol. 251:550-562(1995).
 RN [11]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RC STRAIN-OGAWA 41 / CLASSICAL BIOTYPE;
 RX MEDLINE=97376625; PubMed=9232653;
 RA Merritt E.A., Sarfaty S., Jobling M.G., Chang T., Holmes R.K.,
 RA Hirst T.R., Hol W.G.J.;
 RT "Structural studies of receptor binding by cholera toxin mutants.";
 RL Protein Sci. 6:1516-1528(1997).
 CC -1- FUNCTION: THE BETA CHAIN AGGREGATE (B SUBUNIT) IS INVOLVED IN
 CC BINDING TO CELL MEMBRANES.
 CC -1- SUBUNIT: CONTAINS 3 KINDS OF CHAINS. AN ALPHA AND A GAMMA CHAIN
 CC (FROM THE SAME PRECURSOR MOLECULE), LINKED BY AN INTERCHAIN
 CC DISULFIDE BOND, ASSOCIATE NONCOVALENTLY WITH AN AGGREGATE OF 4 TO
 CC 6 BETA CHAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X00171; CAA24996.1; -
 DR EMBL; K01170; AAA27573.1; -
 DR EMBL; D30053; BAA06291.1; -
 DR EMBL; X58786; CAA1593.1; -
 DR EMBL; X76380; CAA53973.1; -
 DR EMBL; X76391; CAA53976.1; -
 DR EMBL; AE004224; AAF94613.1; -
 DR PIR; A01819; XVCB.
 DR PIR; A05130; A05130.
 DR PIR; S14624; S14624.
 DR PDB; 2CHB; 03-DEC-97.
 DR PDB; 3CHB; 12-AUG-98.
 DR PDB; 1CHP; 08-MAR-96.
 DR PDB; 1CHO; 08-MAR-96.
 DR PDB; 1FGB; 23-DEC-96.
 DR PDB; 1XRB; 01-APR-97.
 DR PDB; 1XTC; 01-AUG-96.

DR PDB; 1CT1; 15-OCT-97.
 DR TIGR; VC1456; -
 DR InterPro: IPR001835; Enterotoxin_B.
 DR Pfam: PF01376; Enterotoxin_B.1.
 DR PRINTS: PR00772; ENTEROTOXINB.
 DR ProDom: PD012805; Enterotoxin_B.1.
 KW Membrane; Enterotoxin; Signal; 3D-structure; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 124 CHOLERA ENTEROTOXIN, BETA CHAIN.
 FT DISULFID 30 107
 FT CONFLICT 33 33 Y -> S (IN REF. 2).
 FT CONFLICT 39 39 Y -> H (IN REF. 7 AND 8).
 FT CONFLICT 43 43 D -> N (IN REF. 7 AND 8).
 FT CONFLICT 68 68 I -> T (IN REF. 7 AND 8).
 FT CONFLICT 70 70 Q -> E (IN REF. 8).
 FT CONFLICT 75 75 G -> S (IN REF. 2).
 FT CONFLICT 91 91 D -> N (IN REF. 7 AND 8).
 FT HELIX 26 30
 FT TURN 31 32
 FT TURN 31 32
 FT TURN 34 35
 FT STRAND 36 44
 FT STRAND 47 51
 FT TURN 54 55
 FT STRAND 58 62
 FT TURN 64 65
 FT STRAND 68 71
 FT TURN 76 77
 FT TURN 80 99
 FT HELIX 100 100
 FT TURN 100 100
 FT STRAND 102 109
 FT STRAND 115 123
 SQ SEQUENCE 124 AA; 13957 MW; 9AA393E3EA8E3EBF CRC64;

Query Match 100.0%; Score 62; DB 1; Length 124;
 Best Local Similarity 100.0%; Pred. No. 6e-05;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VEVPGSQHDSQ 12
 DB 71 VEVPGSQHDSQ 82

RESULT 2
 ID EMBL ECOLI STANDARD; PRT; 124 AA.
 AC P13811;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Heat-labile enterotoxin B chain precursor (LT-B, human) (LTH-B).
 GN ETLB OR LTBP.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OC NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ISOLATE H74-114;
 RX MEDLINE=85156481; PubMed=3884513;
 RA Leong J., Vinal A.C., Dallas W.S.;
 RT "Nucleotide sequence comparison between heat-labile toxin B-subunit
 RT cistrons from Escherichia coli of human and porcine origin.";
 RL Infect. Immun. 48:73-77(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ISOLATE H10407;
 RX MEDLINE=83114628; PubMed=6759877;
 RA Yamamoto T., Tamura T.A., Yokota T., Takano T.;
 RT "Overlapping genes in the heat-labile enterotoxin operon originating
 RT from Escherichia coli human strain.";
 RL Mol. Gen. Genet. 188:356-359(1982).
 RN [3]


```

RP SEQUENCE FROM N.A.
RC STRAIN-ISOLATE H10407;
RX MEDLINE-93252225; PubMed-8486242;
RA Inoue T., Tsuji T., Koto M., Imamura S., Miyama A.;
RT "Amino acid sequence of heat-labile enterotoxin from chicken
RT enterotoxigenic Escherichia coli is identical to that of human strain
RT H 10407."
RL FEMS Microbiol. Lett. 108:157-161(1993).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-ETEC LT 87;
RA Germani Y., Desperrier J.M.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE-95349400; PubMed-7623669;
RA Domenighini M., Plaza M., Jobling M.G., Holmes R.K., Rappuoli R.;
RT "Identification of errors among database sequence entries and
RT comparison of correct amino acid sequences for the heat-labile
RT enterotoxins of Escherichia coli and Vibrio cholerae."
RL Mol. Microbiol. 13:1165-1167(1995).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (3.04 ANGSTROMS).
RX MEDLINE-99185101; PubMed-10085117;
RA Matkovic-Galogovic D., Iloregian A., D'Ancunto M.R., Battistutta R.,
RA Tossi A., Palu G., Zanotti G.;
RT "Crystal structure of the B subunit of Escherichia coli heat-labile
RT enterotoxin carrying peptides with anti-herpes simplex virus type 1
RT activity."
RL J. Biol. Chem. 274:8764-8769(1999).
CC -1- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
CC -1- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M17874; AAA98064.1; -
DR EMBL: J01646; AAB02982.1; -
DR EMBL: S60731; AAC60441.1; -
DR EMBL: X83966; CAA58800.1; -
DR PDB: 1LTR; 23-MAR-99.
DR InterPro: IPR001835; Enterotoxin_B.
DR Pfam: PF01376; Enterotoxin_B.1.
DR PRINTS: PR00772; ENTEROTOXINB.
DR ProDom: PD012805; Enterotoxin_B.1.
DR Enterotoxin; Signal; 3D-structure.
FT SIGNAL 1 21
FT CHAIN 22 124 HEAT-LABILE ENTEROTOXIN B CHAIN.
FT DISULFID 30 107
FT VARIANT 6 6 F -> C (IN ISOLATE H10407).
FT VARIANT 20 20 H -> Y (IN ISOLATE H10407).
FT VARIANT 34 34 H -> R (IN ISOLATE H10407).
SQ SEQUENCE 124 AA; 14027 MW; E9F7F7C7B9D3BC47 CRC64;

Query Match 100.0%; Score 62; DB 1; Length 124;
Best local similarity 100.0%; Pred. No. 6e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSQHDSQ 12
DB 71 VEVPGSQHDSQ 82

RESULT 3
ELBP_ECOLI STANDARD; PRT; 124 AA.
ID ELBP_ECOLI

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M17873; AAA98065.1; -
DR EMBL: M15363; AAA24792.1; -
DR EMBL: M17101; AAA23973.1; -
DR PIR: A01820; QLECB.
DR PIR: B26946; QLECB.
DR PDB: 1LTA; 31-JAN-94.
DR PDB: 1LTB; 31-JAN-94.
DR PDB: 1LTG; 15-SEP-95.
DR PDB: 1LTI; 17-AUG-96.
DR PDB: 1LTS; 31-JAN-94.
DR PDB: 1LTT; 31-JAN-94.
DR PDB: 1LTK; 07-JUL-97.
DR PDB: 1LT4; 16-JUN-97.
DR PDB: 1LTF; 03-DEC-97.
DR PDB: 1LT6; 03-DEC-97.
DR PDB: 1HTL; 20-APR-95.
DR InterPro: IPR001835; Enterotoxin_B.
DR Pfam: PF01376; Enterotoxin_B.1.
DR PRINTS: PR00772; Enterotoxin_B.
DR PRODOM: PD012805; Enterotoxin_B.1.
DR Enterotoxin; Signal; 3D-structure.
KW SIGNAL
FT CHAIN 1 21
FT DISULFID 22 124 HEAT-LABILE ENTEROTOXIN B CHAIN.
FT HELIX 30 107
FT TURN 31 32
FT STRAND 36 43
FT TURN 47 51
FT STRAND 54 55
FT TURN 58 62
FT STRAND 64 65
FT TURN 68 71
FT TURN 76 77
FT HELIX 80 98
FT TURN 99 100
FT STRAND 103 109
FT STRAND 115 123
SQ SEQUENCE 124 AA: 1413 MW: 6DB7DE58395EA70D CRC64;

Query Match 100.0%; Score 62; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 6e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEPGSOHDSQ 12
DB 71 VEPGSOHDSQ 82

RESULT 4
DCMB_MOOTH STANDARD: PRT; 674 AA.
AC P27989;
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Carbon monoxide dehydrogenase beta subunit (EC 1.2.99.2) (CODH)
OS Moorella thermoacetica (Clostridium thermoacetum).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Moorella group;
CC Moorella.
OX NCBI_TaxID=1525;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92084676; PubMed=1748656;
RA Morton T.A., Rungquist J.A., Ragsdale S.W., Shanmugasundaram T.,
RA Wood H.G., Ljungdahl L.G.;
RT "The primary structure of the subunits of carbon monoxide
dehydrogenase/acetyl-CoA synthase from Clostridium thermoacetum.";

RL J. Biol. Chem. 266:23824-23828(1991).
CC -I- FUNCTION: CATALYZES THE INTERCONVERSION OF CO AND CO2 AND THE
CC SYNTHESIS OF ACETYL-COENZYME A FROM THE METHYLATED CORRIDOID/IRON
CC SULEFUR PROTEIN, CO, AND COENZYME A.
CC -I- CATALYTIC ACTIVITY: CO + H(2)O + acceptor = CO(2) + reduced
CC acceptor.
CC -I- COFACTOR: NICKEL.
CC -I- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS.
CC -I- SIMILARITY: TO R.RUBRUM CARBON MONOXIDE DEHYDROGENASE.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
DR EMBL: M62727; AAA23228.1; -
DR PIR: A41670; A41670.
KW Oxidoreductase; Nickel; Iron-sulfur; Electron transport.
FT METAL 68 68 IRON-SULEFUR (BY SIMILARITY).
FT METAL 71 71 IRON-SULEFUR (BY SIMILARITY).
FT METAL 90 90 IRON-SULEFUR (BY SIMILARITY).
SQ SEQUENCE 674 AA: 72924 MW: 34BA3D816C25F9FC CRC64;

Query Match 61.3%; Score 38; DB 1; Length 674;
Best Local Similarity 54.5%; Pred. No. 16;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 EVPGSOHDSQ 12
DB 379 KIPGAYHIDYQ 389

RESULT 5
AROC_YEAST STANDARD: PRT; 376 AA.
AC P28777;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Chorismate synthase (EC 4.6.1.4) (5-enolpyruvylshikimate-3-phosphate
DE phospholase).
GN ARO2 OR YGL148W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=92114793; PubMed=1837329;
RX Jones D.G.L., Reusser U., Braus G.H.;
RT "Molecular cloning, characterization and analysis of the regulation
RT of the ARO2 gene, encoding chorismate synthase, of Saccharomyces
RT cerevisiae.";
RL Mol. Microbiol. 5:2143-2152(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1769;
RX MEDLINE=97197983; PubMed=9046099;
RA Voet M., Defoor E., Verhasselt P., Riles L., Robben J., Volckaert G.;
RT "The sequence of a nearly unclonable 22.8 kb segment on the left arm
RT chromosome VII from Saccharomyces cerevisiae reveals ARO2, RPL9A,
RT TPL, MFI genes and six new open reading frames.";
RL Yeast 13:177-182(1997).
CC -I- CATALYTIC ACTIVITY: 5-O-(1-carboxyvinyl)-3-phosphoshikimate =
CC chorismate + phosphate.
CC -I- COFACTOR: REDUCED FLAVIN.
CC -I- PATHWAY: SEVENTH STEP IN THE BIOSYNTHESIS OF CHORISMATE WITHIN
CC THE BIOSYNTHESIS OF AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).

```
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- INDUCTION: BY AMINO ACID STARVATION.
CC -1- SIMILARITY: BELONGS TO THE CHORISMATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

DR EMBL: X60190: CAA42745.1; -;
DR EMBL: X99960: CAA68214.1; -;
DR EMBL: Z72670: CAA96860.1; -;
DR PIR: S17246: S17246.
DR SGD: S0003116: ARO2.
DR InterPro: IPR000453: Chorismate_synth.
DR Pfam: PF01264: Chorismate_synth; 1.
DR ProDom: PD002941: Chorismate_synth; 1.
DR PROSITE: PS00787: CHORISMATE_SYNTHASE_1; 1.
DR PROSITE: PS00788: CHORISMATE_SYNTHASE_2; 1.
DR PROSITE: PS00789: CHORISMATE_SYNTHASE_3; 1.
KM Lyase: Aromatic amino acid biosynthesis
SQ SEQUENCE 376 AA; 40838 MW; AF3AF65605B91B8E CRC64;

Query Match 59.7%; Score 37; DB 1; Length 376;
Best Local Similarity 70.0%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 VEPGSOQHD 10
1 | | | | | : | |
Db 264 VSPGSKHND 273

RESULT 6
YBSA_THEMEA STANDARD; PRT; 240 AA.
AC P58009;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein TM1158.1.
GN TM1158.1.
OS Thermotoga maritima.
OC Bacteria: Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linner K.D., Garrett M.M.,
RA Stewart A.M., Sutton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
RP IDENTIFICATION.
RA Medigue C., Bocs S.;
RT Unpublished observations (APR-2001).

```
CC -----
CC EMBL: AE001773; -; NOT_ANNOTATED_CDS.
CC TIGR: TM1158.1; -.
CC KW Hypothetical protein; Complete proteome.
CC SEQUENCE 240 AA; 27773 MW; 1BEF66C1C8BD2700 CRC64;
```

Query Match 58.1%; Score 36; DB 1; Length 240;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 VPGSOHDSQ 12
1 | | | | | : | |
Db 226 VPGSEHLEK 235

RESULT 7
GUDH_BACSU STANDARD; PRT; 455 AA.
ID GUDH_BACSU
AC P4238;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable glucarate dehydratase (EC 4.2.1.40) (GDH) (GLUD).
GN GUDH.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95219079; PubMed=7704254;
RA Ogawa K.-I., Akagawa E., Nakamura K., Yamane K.;
RT "Determination of a 21548 bp nucleotide sequence around the 24
RT degrees region of the Bacillus subtilis chromosome.";
RL Microbiology 141:269-275(1995).
CC -1- FUNCTION: CATALYZES THE DEHYDRATION OF GLUCARATE TO 5-KETO-4-
CC DEOXY-D-GLUCARATE (5-KDGLUC) (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: D-glucarate -> 5-dehydro-4-deoxy-D-glucarate +
CC H(2)O.
CC -1- PATHWAY: FIRST STEP IN GLUCARATE CATABOLISM.
CC -1- SIMILARITY: BELONGS TO THE MANDELATE RACEMASE / MUCONATE
CC LACTONIZING ENZYME FAMILY. GLUCD SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: D30808: BAA06470.1; -;
DR EMBL: Z99105: CAB12043.1; -;
DR HSSP: P42206: 1BQG.
DR Subtilisin, BG11161; GUDH.
DR InterPro: IPR001354; MR_MLE.
DR Pfam: PF01188; MR_MLE; 1.
DR Pfam: PF02746; MR_MLE_N; 1.
KM Lyase: Complete proteome.
SQ SEQUENCE 455 AA; 50782 MW; 3238486007698C2A CRC64;

Query Match 58.1%; Score 36; DB 1; Length 455;
Best Local Similarity 75.0%; Pred. No. 26;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVPGSOQHT 9
1 | | | | | : | |
Db 64 EVPGSEHT 71

RESULT 8
BGAL_THEFT
ID BGAL_THEFT STANDARD: PRT: 743 AA.
AC P77989;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Beta-galactosidase (EC 3.2.1.23) (Lactase).
GN LAC2 OR LACA.
OS Thermomonas aerobacter ethanolicus (Clostridium thermohydrosulfuricum).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermomonas aerobacter group; Thermomonas aerobacter.
OX NCBI_TaxID=1757;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33223 / 39E;
RA Zverlov V.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC galactose residues in beta-D-galactosides.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y08557; CAA69850.1; -
DR InterPro: IPR001649; Glyco_hydro.2.
DR Pfam: PF00703; Glyco_hydro.2; 1.
DR Pfam: PF02836; Glyco_hydro.2_C; 1.
DR Pfam: PF02837; Glyco_hydro.2_N; 1.
DR PROSITE: PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR PROSITE: PS00608; GLYCOSYL_HYDROL_F2_2; 1.
KM Hydrolyase; Glycosidase.
FT ACT_SITE 388 388 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 453 453 NOCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 743 AA; 85796 MW; FE01FE517E51DFC CRC64;
QY 2 EYVGSQHDSD 12
DB 346 EIRGMQHDGDE 356
Query Match 58.1%; Score 36; DB 1; Length 743;
Best Local Similarity 54.5%; Pred. No. 44;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
RESULT 9
ST14_HUMAN STANDARD: PRT: 855 AA.
AC Q9Y5T6; Q9H3S0; Q9HCA3; Q9HS01; Q9HB36;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Matrilysin) (Membrane-
DE type serine protease 1) (MT-Sp1) (Prostasin) (Serine protease TADG-15)
DE (Tumor associated differentially-expressed gene-15 protein).
GN ST14 OR PRSS14 OR SMC19 OR TADG15.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RP MEDLINE=99303581; PubMed=10373424;
RA Lin C.Y., Anders J., Johnson M., Sang O.A., Dickson R.B.;
RT "Molecular cloning of cDNA for matrilysin, a matrix-degrading serine
RT protease with trypsin-like activity.";

RL J. Biol. Chem. 274:18231-18236(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99432178; PubMed=10500122;
RA Takeuchi T., Shuman M.A., Craik C.S.;
RT "Reverse biochemistry: Use of macromolecular protease inhibitors to
RT dissect complex biological processes and identify a membrane-type
RT serine protease in epithelial cancer and normal tissue.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Yamaguchi N., Mitsui S.;
RT "Molecular cloning of a novel transmembrane serine protease expressed
RT in human prostate.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Tanimoto H., Underwood L.J., Wang Y., Shigemasa K., Parmley T.H.,
O'Brien T.J.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 327-855 FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 340-664 FROM N.A.
RA Cao J., Fan W., Zheng S.;
RT "Genomic analysis of a novel human serine protease SMC19.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP CHARACTERIZATION.
RC TISSUE=Milk;
RX MEDLINE=99303582; PubMed=10373425;
RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;
RT "Purification and characterization of a complex containing matrilysin
RT and a kunitz-type serine protease inhibitor from human milk.";
RL J. Biol. Chem. 274:18237-18242(1999).
CC -1- FUNCTION: DEGRADATES EXTRACELLULAR MATRIX. PROPOSED TO PLAY A ROLE
CC IN BREAST CANCER INVASION AND METASTASIS. EXHIBITS TRYPSIN-LIKE
CC ACTIVITY AS DEFINED BY CLEAVAGE OF SYNTHETIC SUBSTRATES WITH ARG
CC OR LYS AS THE P1 SITE.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF118224; AADA2765.2; -
DR EMBL: AF133086; AAF00109.1; -
DR EMBL: AB030036; BAB20376.1; -
DR EMBL: AF057145; AAG15395.1; -
DR EMBL: BC005826; AAH05826.1; -
DR EMBL: AF283256; AAG13949.1; -
DR HSPS: P00763; IDPO.
DR MEROPS: S01.302; -
DR InterPro: IPR000859; CUB.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR002172; LDL_recept_A.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00087; Idl_recept_a; 4.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.

CC EMBL: M57500; AAC64900.1; -
 DR PIR: JN0143; JN0143.
 DR HSSP: P00437; 3PCD.
 DR InterPro: IPR000627; Dioxigenase.
 DR Pfam: PF00775; Dioxigenase; 1.
 DR PROSITE: PS00083; INTRADIOL_DIOXYGENAS; 1.
 KW Aromatic hydrocarbons catabolism; Oxidoreductase; Dioxigenase; Iron;
 KW Plasmid.
 FT METAL 164 164 IRON (BY SIMILARITY).
 FT METAL 198 198 IRON (BY SIMILARITY).
 FT METAL 222 222 IRON (BY SIMILARITY).
 FT METAL 224 224 IRON (BY SIMILARITY).
 SQ SEQUENCE 302 AA; 33362 MW; A86F17E68D1EAC3A CRC64;

Query Match 56.5%; Score 35; DB 1; Length 302;
 Best Local Similarity 41.7%; Pred. No. 25;
 Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Oy 1 VEPGSOHIDSQ 12
 Db 227 ISAPGHQHLTQ 238

RESULT 12
 ID YP68_CAEEL STANDARD; PRT; 313 AA.
 AC 009217;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 37.0 kDa protein B0495.8 in chromosome II.
 GN B0495.8.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Kirsten J;
 RL Submitted (FEB-1995) to the EMBL/Genbank/DBI databases.
 CC -1- SIMILARITY: SOME, TO YEAST YD1087C AND S. POMBE SPEC16A11.13.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U21317; AAA62527.1; -
 DR WormPep: B0495.8; CE01766.
 KW Hypothetical protein.
 SQ SEQUENCE 313 AA; 36977 MW; 000D2327621BPE0D CRC64;

Query Match 56.5%; Score 35; DB 1; Length 313;
 Best Local Similarity 45.5%; Pred. No. 26;
 Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Oy 2 EYPGSOHIDSQ 12
 Db 11 QLMGSHVDNK 21

RESULT 13
 YM18_PSEAE STANDARD; PRT; 371 AA.
 ID YM18_PSEAE
 AC 001609; 0911P9;
 DT 01-FEB-1995 (Rel. 31, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein PA2218.
 GN PA2218.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01 / H103;
 RX MEDLINE=93051258; PubMed=1427017;
 RA Huang H., Stenel R.J., Bellido F., Rawling E., Hancock R.E.W.;
 RT "Analysis of two gene regions involved in the expression of the
 RT imipenem-specific, outer membrane porin protein OprD of Pseudomonas
 RT aeruginosa."
 RL FEMS Microbiol. Lett. 76:267-274(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen."
 RL Nature 406:959-964(2000).
 CC -1- SIMILARITY: TO E.COLI YC9Y.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: Z14064; CAA78447.1; -
 DR EMBL: AF004648; AAG05606.1; ALT_INIT.
 DR PIR: S23861; S23861.
 KW Hypothetical protein; Complete proteome.
 FT CONFLICT 16 R -> P (IN REF. 1).
 FT CONFLICT 73 R -> P (IN REF. 1).
 FT CONFLICT 261 A -> G (IN REF. 1).
 SQ SEQUENCE 371 AA; 40840 MW; D7EB0CCAC95A7CF6 CRC64;

Query Match 56.5%; Score 35; DB 1; Length 371;
 Best Local Similarity 62.5%; Pred. No. 32;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 3 VPGSOHID 10
 Db 343 VPGASHVD 350

RESULT 14
 TYRO_ASPO STANDARD; PRT; 539 AA.
 ID TYRO_ASPO
 AC 000234;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Tyrosinase (EC 1.14.18.1) (Monophenol monooxygenase).
 GN MEIO.
 OS Aspergillus oryzae.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5062;
 RN [1]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 3, 2002, 09:06:02 ; Search time 74.17 Seconds
(without alignments)
27.989 Million cell updates/sec

Title: US-09-786-648-3
Perfect score: 62
Sequence: 1 VEVPSQHDSQ 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_19:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	100.0	103	2	Q9R646
2	62	100.0	124	2	Q57193
3	62	100.0	124	2	Q9RPI5
4	62	100.0	124	2	Q56635
5	62	100.0	124	2	Q83V32
6	62	100.0	124	2	Q94M01
7	43	69.4	395	5	Q9NMD5
8	42	67.7	802	10	Q947M6
9	42	67.7	1109	10	Q94HM6
10	42	67.7	1352	10	Q94EB4
11	42	67.7	2535	10	Q948C7
12	40	64.5	242	2	Q86582
13	39	64.5	254	4	Q31335
14	39	62.9	192	4	Q9NPO6
15	39	62.9	201	2	Q9RJZ8
16	39	62.9	233	10	Q94KK7

17	39	62.9	260	10	Q9MA16	Q9MA16 arabidopsis
18	39	62.9	272	4	Q9UG04	Q9UG04 homo sapien
19	39	62.9	353	16	Q9RVJ7	Q9RVJ7 deinococcus
20	39	62.9	412	4	Q9Y2B3	Q9Y2B3 homo sapien
21	39	62.9	428	10	Q9LGM2	Q9LGM2 oryza sativ
22	39	62.9	1076	10	Q94I82	Q94I82 oryza sativ
23	39	62.9	1242	13	Q90Y57	Q90Y57 brachydanio
24	38.5	62.1	565	10	Q22511	Q22511 vitis vinif
25	37	59.7	779	4	Q96HT2	Q96HT2 homo sapien
26	37	59.7	331	5	Q18391	Q18391 drosophila
27	37	59.7	354	3	Q14347	Q14347 schizosacch
28	37	59.7	367	10	Q80418	Q80418 nicotiana t
29	37	59.7	423	10	Q9C531	Q9C531 arabidopsis
30	37	59.7	457	10	Q949W1	Q949W1 arabidopsis
31	37	59.7	508	4	Q99887	Q99887 homo sapien
32	37	59.7	550	10	Q04636	Q04636 arabidopsis
33	37	59.7	899	3	Q96W69	Q96W69 candida alb
34	37	59.7	899	3	Q96W68	Q96W68 candida alb
35	37	59.7	1328	10	Q9AU17	Q9AU17 lycopersico
36	36	58.1	373	2	Q9XD79	Q9XD79 streptomyce
37	36	58.1	382	5	Q9V7M7	Q9V7M7 drosophila
38	36	58.1	389	5	Q9U0W5	Q9U0W5 leishmania
39	36	58.1	527	12	Q9E226	Q9E226 helioverpa
40	36	58.1	528	12	Q99G79	Q99G79 heliocoverp
41	36	58.1	551	10	Q9S079	Q9S079 pinus taeda
42	36	58.1	745	2	Q93IM0	Q93IM0 thermoanaer
43	36	58.1	755	16	Q92XF7	Q92XF7 thizobium m
44	36	58.1	832	12	Q9D0H5	Q9D0H5 hepatitis b
45	36	58.1	832	12	Q9D0H1	Q9D0H1 hepatitis b

ALIGNMENTS

RESULT 1	
Q9R646	PRELIMINARY: PRT: 103 AA.
AC Q9R646:	
DT 01-MAY-2000 (TREMBLrel. 13, Created)	
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)	
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)	
DE CHOLERA-LIKE ENTEROTOXIN B SUBUNIT.	
OS Vibrio cholerae.	
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.	
OX NCBI_maxID=666;	
RN [1]	
RP SEQUENCE.	
RX MEDLINE=95303036; PubMed=7783690;	
RA Nakashima K., Eguchi Y., Nakasone N.;	
RT "Characterization of an enterotoxin produced by Vibrio cholerae	
RT O139."	
RL Microbiol. Immunol. 39:87-94(1995).	
DR HSSP; P01556; 1XTC.	
DR InterPro: IPR001835; Enterotoxin_B.	
DR Pfam: PF01376; Enterotoxin_B. 1.	
DR PRINTS: PR00772; ENTEROTOXINB.	
DR PRODOM: PD012805; Enterotoxin_B. 1.	
SQ SEQUENCE 103 AA; 11645 MW; 992A05C077F2F70E CRC64;	
Query Match	100.0% Score 62; DB 2; Length 103;
Best Local Similarity	100.0% Pred. NO. 0.00019;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY 1 VEVPSQHDSQ 12	
DB 50 VEVPSQHDSQ 61	
RESULT 2	
Q57193	PRELIMINARY: PRT: 124 AA.
ID Q57193:	
AC Q57193:	

DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
 DE CHOLERA TOXIN B PROTEIN (CTB) PRECURSOR (CTB).
 GN CTXB.
 OS *Vibrio cholerae*.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; *Vibrio*.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLASSICAL STRAIN 569B;
 RX MEDLINE=91355224; PubMed=1883840;
 RA Dams E., De Wolf M., Dierick W.;
 RT "Nucleotide sequence analysis of the CT operon of the *Vibrio cholerae*
 RL classical strain 569B."
 RL Biochim. Biophys. Acta 1090:139-141(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLASSICAL BIOTYPE 569B;
 RA Shi C., Cao C., Zhang J., Ma Q.;
 RL Chin. Biochem. J. 9:395-399(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLASSICAL BIOTYPE 569B;
 RA Xu L.;
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X58785; CAA41591.1; -
 DR EMBL: U25679; AAC34728.1; -
 DR EMBL: A00931; CAA00098.1; -
 DR HSSP: P01556; 2CHB.
 DR InterPro: IPR001835; Enterotoxin_B.
 DR Pfam: PF01376; Enterotoxin_B; 1.
 DR PRINTS: PR00772; ENTEROTOXINB.
 DR ProDom: PD012805; Enterotoxin_B; 1.
 KW Signal.
 FT SIGNAL. 1 21 POTENTIAL.
 FT CHAIN 22 124 CHOLERA TOXIN B PROTEIN (CTB).
 SQ SEQUENCE 124 AA; 13919 MW; D6BF83FEF7924EA3 CRC64;

Query Match 100.0%; Score 62; DB 2; Length 124;
 Best Local Similarity 100.0%; Pred. No. 0.00023;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSQHIDSQ 12
 |||||
 DB 71 VEVPGSQHIDSQ 82

RESULT 3
 ID 09RP15 PRELIMINARY; PRT; 124 AA.
 AC 09RP15:
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
 DE CHOLERA ENTEROTOXIN B-SUBUNIT.
 GN CTXB.
 OS *Vibrio cholerae*.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; *Vibrio*.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KNH002;
 RA Shin H.J., Park Y.C., Kim Y.C.;
 RT "Cloning and nucleotide sequence analysis of the virulence gene
 RT cassette from *Vibrio cholerae* KNH002 isolated in Korea."
 RL Misaimurhag Hoji 35:205-210(1999).
 DR EMBL: AF175708; AAD51360.1; -
 DR HSSP: P01556; 2CHB.
 DR InterPro: IPR001835; Enterotoxin_B.
 DR Pfam: PF01376; Enterotoxin_B; 1.
 DR PRINTS: PR00772; ENTEROTOXINB.

DR ProDom: PD012805; Enterotoxin_B; 1.
 SQ SEQUENCE 124 AA; 13905 MW; 23BF83FEF793E5B9 CRC64;

Query Match 100.0%; Score 62; DB 2; Length 124;
 Best Local Similarity 100.0%; Pred. No. 0.00023;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSQHIDSQ 12
 |||||
 DB 71 VEVPGSQHIDSQ 82

RESULT 4
 ID 056635 PRELIMINARY; PRT; 124 AA.
 AC 056635:
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
 DE CHOLERA TOXIN PRECURSOR.
 GN CTXB.
 OS *Vibrio cholerae*.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; *Vibrio*.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=57;
 RA Honda T.;
 RA Yamamoto K., Do V.G., Xu M., Tida T., Miwatani T., Albert M.J.,
 RT "Comparison of cholera toxin genes (ctxAB) of non-O1 *Vibrio cholerae*
 RT strains 854 (O139-bengal) and 57 (O37) from two outbreaks."
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
 DR EMBL: D30052; BAA06289.1; -
 DR HSSP: P01556; 2CHB.
 DR InterPro: IPR001835; Enterotoxin_B.
 DR Pfam: PF01376; Enterotoxin_B; 1.
 DR PRINTS: PR00772; ENTEROTOXINB.
 DR ProDom: PD012805; Enterotoxin_B; 1.
 KW Signal.
 FT SIGNAL. 1 21 POTENTIAL.
 FT SEQUENCE 124 AA; 13871 MW; 3f87B2F297953179 CRC64;

Query Match 100.0%; Score 62; DB 2; Length 124;
 Best Local Similarity 100.0%; Pred. No. 0.00023;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSQHIDSQ 12
 |||||
 DB 71 VEVPGSQHIDSQ 82

RESULT 5
 ID 093V32 PRELIMINARY; PRT; 124 AA.
 AC 093V32:
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE HEAT-LABILE ENTEROTOXIN B SUBUNIT.
 GN LTH B SUBUNIT.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OX *Escherichia*.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1032 (ENTEROTOXIGENIC);
 RA Komase K.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN-1032 (ENTEROTOXIGENIC);
 RA MEDLINE=95091056; PubMed=7998417;
 RX Tamura S., Asanuma H., Tomita T., Komase K., Kawahara K., Danbara H.,
 RA Hatori N., Watanabe K., Suzuki Y., Nagamine T., Aizawa C., Oya A.,
 RA Kurata T.;
 RT "Escherichia coli heat-labile enterotoxin B subunits supplemented with
 RT a trace amount of the holotoxin as an adjuvant for nasal influenza
 RT vaccine.";
 RL Vaccine 12:1083-1089(1994).
 DR EMBL, AB011677; BAA25726.1; -
 SQ SEQUENCE 124 AA; 14028 MW; 5346BD3BB32354C2 CRC64;

Query Match 100.0%; Score 62; DB 2; Length 124;
 Best Local Similarity 100.0%; Pred. NO. 0.00023;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEPGSOHIDSQ 12
 Db 71 VEPGSOHIDSQ 82

RESULT 6
 O94M01 PRELIMINARY; PRT; 124 AA.
 AC O94M01;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE CTXB.
 GN CTXB.
 OS Vibrio phage CTX.
 OC Viruses; ssDNA viruses; Inoviridae; Inovirus.
 OX NCBI_TaxID=141904;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bhattechariyaa T., Nandy R.K., Nair G.B.;
 RT "The entire core region of the ctx-phi (ctx-prophage) in VCE 232, an
 RT environmental strain of V. cholerae.";
 RL Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL, AF414369; AL009682.1; -
 SQ SEQUENCE 124 AA; 13884 MW; 8D4C251591B16891 CRC64;

Query Match 100.0%; Score 62; DB 9; Length 124;
 Best Local Similarity 100.0%; Pred. NO. 0.00023;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEPGSOHIDSQ 12
 Db 71 VEPGSOHIDSQ 82

RESULT 7
 O9NKM5 PRELIMINARY; PRT; 395 AA.
 AC O9NKM5;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE HYPOHETICAL 43.6 KDA PROTEIN.
 GN BG:DS01514.3.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN-Y, CN BW SP;
 RX MEDLINE=99403001; PubMed=10471707;
 RA Ashburner M., Misra S., Roote J., Lewis S.E., Blazer R., Davis T.,
 Doyle C., Galle R., George R., Harris N., Hartnell G., Harvey D.,

RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
 RA Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitehead K.,
 RA Celniker S., Rubin G.M.;
 RT "An exploration of the sequence of a 2.9-Mb region of the genome of
 RT Drosophila melanogaster: the Adh region.";
 RL Genetics 153:179-219(1999).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN-Y, CN BW SP;
 RA Celniker S.E., Agbayan A., Arcaina T.T., Baxter E., Blazer R.G.,
 RA Butenoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
 RA Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
 RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
 RA Lewis S., Li P., Lomoton M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
 RA Nixon K., Paclab J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
 RA Sehn H., Snir E., Svirskaas R.R., Wan K.H., Weinburg T., Zhang R.,
 RA Zieran L.L., Rubin G.M.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL, AE003408; AAF44847.1; -
 DR FlyBase, FBgn0028907; BG:DS01514.3.
 KW Hypothetical protein.
 SQ SEQUENCE 395 AA; 43561 MW; AE4F1CC4AD3DA73 CRC64;

Query Match 69.4%; Score 43; DB 5; Length 395;
 Best Local Similarity 72.7%; Pred. NO. 4.2;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 VEPGSOHIDS 11
 Db 330 VSPGSTHIDA 340

RESULT 8
 O94TW6 PRELIMINARY; PRT; 802 AA.
 AC O94TW6;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE PUTATIVE TNP-LIKE TRANSPOSABLE ELEMENT.
 GN OSJNBA0029P06.9.
 OS Oryza sativa (Rice)
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Euphorbiaceae; Oryzae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wing R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
 RA Saeki C., Henry D., Oates R., Simmons J.;
 RT "Rice Genomic Sequence.";
 RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL, AC090873; AAK98653.1; -
 SQ SEQUENCE 802 AA; 92934 MW; 014DE54BA1CD108 CRC64;

Query Match 67.7%; Score 42; DB 10; Length 802;
 Best Local Similarity 77.8%; Pred. NO. 14;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 4 PGOHIDSQ 12
 Db 516 PGOHIESE 524

RESULT 9
 O94HM6 PRELIMINARY; PRT; 1109 AA.
 AC O94HM6;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

RL MOL. Microbiol. 21:77-96(1996).
 DR EMBL: AL031184; CAA20190.1; -;
 DR InterPro: IPR002502; Amidase_2.
 DR Pfam: PF01510; Amidase_2; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 242 AA; 26018 MW; 6DD10FF18A2EC544 CRC64;

Query Match
 Best Local Similarity 64.5%; Score 40; DB 2; Length 242;
 Pred. No. 9.3;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 VEPGSOHID 10
 |||||: ||
 Db 204 VEPGTDHTD 213

RESULT 13
 ID 031335 PRELIMINARY; PRT; 254 AA.

AC 031335;
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE APP-DEPENDENT NUCLEASE SUBUNIT A (FRAGMENT).
 CN ADDA.
 OS Bacillus cereus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_Taxid=1396;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 10987;
 RA MEDLINE=99231848; PubMed=10217496;
 RA Ostad O.A., Hegna I., Lindbaeck T., Rishovd A.L., Kolsto A.B.;
 RT "Genome organisation is not conserved between Bacillus cereus and
 RT Bacillus subtilis";
 RL Microbiology 145:621-631(1999).
 DR EMBL: Y11217; CAA72103.1; -;
 DR InterPro: IPR000212; Vrd-helicase.
 DR Pfam: PF00580; Vrd-helicase; 1.
 FT NON_TER 1 1
 FT 254 254
 SQ SEQUENCE 254 AA; 29212 MW; 24392E1133BD992A CRC64;

Query Match
 Best Local Similarity 64.5%; Score 40; DB 2; Length 254;
 Pred. No. 9.8;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 VEPGSOHIDSO 12
 :: ||||| |
 Db 35 IDEPGSOHIRKO 46

RESULT 14
 ID 09NP06 PRELIMINARY; PRT; 192 AA.
 AC 09NP06;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE LfL, LCAT-LIKE LYSOPHOSPHOLIPASE (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Pluvinet R., Estivill X., Escarceller M., Sumoy L.;
 RL submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.

RA Auffray C., Ansgore W., Ballabio A., Estivill X., Gibson K.,
 RA Lehrach H., Poustka A., Lundeberg J.;
 RT "The European IMAGE consortium for integrated molecular analysis of
 RT human gene transcripts."
 RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AL389957; CAB97531.1; -;
 FT NON_TER 1 1
 SQ SEQUENCE 192 AA; 21609 MW; 04A7AE8CB344F213 CRC64;

Query Match
 Best Local Similarity 62.9%; Score 39; DB 4; Length 192;
 Pred. No. 11;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 EVPGSOHID 10
 |||||: ||
 Db 166 ELPGESEHIE 174

RESULT 15
 ID 09RJ28 PRELIMINARY; PRT; 201 AA.

AC 09RJ28;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE PUTATIVE AMIDASE.
 GN SCG11A.03.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_Taxid=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Oliver K., Harris D.;
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinashl H., Hopwood D.A.;
 RT "A set of ordered cosmid and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL MOL. Microbiol. 21:77-96(1996).
 DR EMBL: AL133210; CAB61584.1; -;
 DR InterPro: IPR002502; Amidase_2.
 DR Pfam: PF01510; Amidase_2; 1.
 SQ SEQUENCE 201 AA; 22749 MW; B8EFA77E06A20468 CRC64;

Query Match
 Best Local Similarity 62.9%; Score 39; DB 2; Length 201;
 Pred. No. 12;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 EVPGSOHID 10
 |||||: ||
 Db 176 EVPGSDHTD 184

Search completed: July 3, 2002, 09:06:03
 Job time: 617 sec

XX Disclosure; Page 15; 62pp; English.
PS
XX
CC The invention relates to peptide fragments of the *Escherichia coli* heat
CC labile enterotoxin (Etx) and its closely related homologue, cholera
CC toxin (Ctx) from *Vibrio cholerae* which do not bind to the ubiquitous
CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are
CC composed of one A subunit and five identical B subunits. The A subunit
CC is responsible for toxicity, possessing adenosine diphosphate (ADP)-
CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)
CC facilitate the entry of subunit A into the host cell via the binding and
CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible
CC for some of the effects of Etx and Ctx, it has been found that certain
CC effects of the toxins, such as immunomodulation, are not mediated
CC through GM-1 binding. The peptides of the invention are fragments of the
CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as
CC normal EtxB and CtxB subunits, except that they do not bind or cross link
CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.
CC They may also be used as an inhibitor for toxin-induced diarrhoea.
CC Therefore, the peptides may be used in the production of a composition
CC for treating, preventing and/or modulating a disease associated with an
CC immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463
CC represent preferred peptides of the invention, AAY87460 being
CC particularly preferred.
XX
SQ Sequence 21 AA:

Query Match 100.0%; Score 106; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 5.1e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSKKAI 21
|||
DB 1 gatfgyevpgsqhidsqkai 21

RESULT 2
AAM04857
ID AAM04857 standard; Protein; 103 AA.
XX
AC AAM04857;
XX
DT 21-FEB-1997 (first entry)
XX
XX Synthetic cholera toxin B subunit.
DE
XX
KM Bordetella pertussis; whooping cough; recombinant construct;
KM cholera toxin B subunit; enzyme; antigen; immunogen; allergen;
KM enzyme inhibitor; hormone; lymphokine; immunoglobulin; toxin;
KM structural protein; receptor; heterologous gene; leader; promoter.
XX
OS Synthetic.
XX
PN WO9626282-A1.
XX
PD 29-AUG-1996.
XX
PF 23-FEB-1996; 96WO-CA00107.
XX
PR 23-FEB-1995; 95US-0393334.
XX
PA (CONN-) CONNUGHT LAB LTD.
XX
PI Klein MH, Loosmore SM, Yacoub RK, Zealey GR;
XX
DR WPI; 1996-425088/42.
DR N-PSDB; AAT38038.
XX
XX Recombinant constructs for expressing and opt. secreting proteins in
PT Bordetella - comprise Bordetella promoter coupled to non-Bordetella,
PT esp. cholera B toxin, gene or coupled to non-Bordetella leader and
PT gene of interest

XX
PS Example 1; Figure 1; 61pp; English.
XX
CC Recombinant constructs comprising a promoter functional in
CC Bordetella operatively linked to a heterologous gene or a non-
CC Bordetella leader sequence for secretion of a gene product which may
CC or may not be of Bordetella origin, can be used for the expression
CC in Bordetella of enzymes, antigens, immunogens, allergens, enzyme
CC inhibitors, hormones, lymphokines, immunoglobulins or their
CC fragments, toxins, mammalian proteins, structural proteins or
CC receptors. The Bordetella strains are particularly engineered to
CC express the cholera toxin B subunit (this sequence). The promoters
CC used in the constructs are selected from the Bordetella pertussis
CC tox, fha promoters or the high molecular weight (hmw) outer membrane
CC promoter of non typable Haemophilus influenzae; leaders used in the
CC constructs are selected from the cholera toxin B leader (CTB-L), the
CC pertussis toxin subunit S1 leader (ST-L) and the pertussin pertactin
CC leader (PRN-L); and genes used in the constructs are selected from a
CC novel synthetic cholera toxin B gene (ctb) and the hmw1 and hmw2
CC genes of Haemophilus influenzae.
XX
SQ Sequence 103 AA:

Query Match 100.0%; Score 106; DB 17; Length 103;
Best Local Similarity 100.0%; Pred. No. 3.2e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSKKAI 21
|||
DB 45 gatfgyevpgsqhidsqkai 65

RESULT 3
AAR94939
ID AAR94939 standard; Protein; 103 AA.
XX
AC AAR94939;
XX
DT 31-OCT-1996 (first entry)
XX
XX Heat labile enterotoxin B subunit (LT-B) E.coli.
DE
XX
KM Toxin; subunit; vaccine; transgenic plant; immunogen; antigen;
KM adjuvant; immunisation.
XX
XX
OS *Escherichia coli*.
XX
PN WO9612801-A1.
XX
PD 02-MAY-1996.
XX
PF 24-OCT-1995; 95WO-US13376.
XX
PR 24-OCT-1994; 94US-0328716.
XX
PA (TULA) TULANE EDUCATIONAL FUND.
PA (TEXA) UNIV TEXAS A & M SYSTEM.
XX
PI Arntzen CJ, Clements JD, Haq TA, Mason HS;
XX
DR WPI; 1996-230602/23.
DR N-PSDB; AAT18799, AAT18800.
XX
XX Transgenic plants contg. *E. coli* heat labile enterotoxin subunits
PT used as oral vaccines for animals which consume the plant
XX
PS Disclosure; Page 100-101; 130pp; English.
XX
XX A transgenic plant comprising or expressing a DNA sequence encoding
CC an immunogenic agent can be used as an oral vaccine for animals.
CC The vaccine is administered by the oral consumption of the plant and
CC provides the first known functional method for immunising animals

CC using transgenic plants, where the plants express bacterial antigens
 CC that act as both immunogens and adjuvants. The method provides an
 CC inexpensive production and delivery system for such antigens to
 CC animals. This is the LT-B Escherichia coli toxin subunit and its
 CC coding sequence was used in the construction of such a transgenic
 CC plant. The immunogenic agent preferably comprises the LT-B or CT-B
 CC (cholera toxin B subunit) or optionally LT-A or CT-A.

XX Sequence 103 AA;

Query Match 100.0%; Score 106; DB 17; Length 103;
 Best Local Similarity 100.0%; Pred. No. 3.2e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATQVEVPSQHIDSQKAI 21
 |||
 Db 45 gatfvevpsqhdsgkai 65

RESULT 4

AAW06606 standard; Protein; 103 AA.

XX AAW06606;

DT 06-AUG-1997 (first entry)

XX Cholera toxin B subunit/heat labile enterotoxin B subunit hybrid.

KM heat labile enterotoxin B subunit; hybrid; cholera toxin B subunit; CTB;

KW immunogen; antigen; broad spectrum; cross reactive; diarrhoea;

XX vomiting; food poisoning.

OS Chimeric - Vibrio cholerae.

XX Chimeric - Enterotoxigenic Escherichia coli.

FT key location/Qualifiers

FT Misc-difference 1 /label= substitution

FT /note= "wild-type Thr replaced by Ala"

FT Misc-difference 94 /label= substitution

FT /note= "wild-type His replaced by Asn"

FT Misc-difference 95 /label= substitution

FT /note= "wild-type Ala replaced by Ser"

PN WO9634893-A1.

XX 07-NOV-1996.

XX 02-MAY-1996; 96WO-SE00570.

XX 05-MAY-1995; 95SE-0001682.

XX (HOLM/) HOLMGREN J.

XX (LEBE/) LEBENS M R.

XX Holmgren J, Lebens MR;

XX WPI; 1996-506108/50.

XX N-PSDB; AAT43576.

XX Heat-labile enterotoxin B sub-unit and cholera toxin B sub-unit

XX hybrid protein - opt. fused to immunogenic sequence for use in

XX vaccines against enterotoxin-induced illness

XX Claim 3; Page -: 32pp; English.

XX AAW06606 is a mature cholera toxin B subunit (CTB)/heat labile

XX enterotoxin B subunit (LTB) hybrid molecule. In mature CTB protein,

CC certain amino acids (aa) were replaced with corresponding aa from
 CC heat-labile enterotoxin B subunit (LTB). The specific amino acid
 CC substitutions impart LTB-specific epitope characteristics to
 CC immunogenic mature CTB. The hybrid molecules have increased
 CC cross-reactivity and are suitable for a broad spectrum vaccine to
 CC protect against enterotoxigenic illness. Immunogenic proteins
 CC comprising the hybrid molecules can be used to treat, or in a
 CC vaccine to prevent, enterotoxin-induced illness, e.g. diarrhoea
 CC and vomiting, in humans and animals.

XX Note - This sequence does not appear in the specification, it is

XX a claimed mutant sequence of mature cholera toxin B subunit (see

XX AAW06605).

RESULT 5

AAW06607 standard; Protein; 103 AA.

XX AAW06607;

DT 06-AUG-1997 (first entry)

XX Cholera toxin B subunit/heat labile enterotoxin B subunit hybrid.

KM heat labile enterotoxin B subunit; hybrid; cholera toxin B subunit; CTB;

KW immunogen; antigen; broad spectrum; cross reactive; diarrhoea;

XX vomiting; food poisoning.

OS Chimeric - Vibrio cholerae.

XX Chimeric - Enterotoxigenic Escherichia coli.

FT key location/Qualifiers

FT Misc-difference 1..25 /label= substitution

FT /note= "the first 25 amino acids of mature

FT wild-type cholera toxin B subunit are

FT replaced with the first 25 amino acids

FT of mature enterotoxin B subunit"

PN WO9634893-A1.

XX 07-NOV-1996.

XX 02-MAY-1996; 96WO-SE00570.

XX 05-MAY-1995; 95SE-0001682.

XX (HOLM/) HOLMGREN J.

XX (LEBE/) LEBENS M R.

XX Holmgren J, Lebens MR;

XX WPI; 1996-506108/50.

XX N-PSDB; AAT43577.

XX Heat-labile enterotoxin B sub-unit and cholera toxin B sub-unit

XX hybrid protein - opt. fused to immunogenic sequence for use in

XX vaccines against enterotoxin-induced illness

XX Claim 4; Page -: 32pp; English.

AAW06607 is a mature cholera toxin B subunit (CTB)/heat labile enterotoxin B subunit (LTB) hybrid molecule. In mature CTB protein, certain amino acids (aa) were replaced with corresponding aa from heat-labile enterotoxin B subunit (LTB). The specific amino acid substitutions impart LTB-specific epitope characteristics to immunogenic mature CTB. The hybrid molecules have increased cross-reactivity and are suitable for a broad spectrum vaccine to protect against enterotoxigenic illness. Immunogenic proteins comprising the hybrid molecules can be used to treat, or in a vaccine to prevent, enterotoxin-induced illness, e.g. diarrhoea and vomiting, in humans and animals.

Note - This sequence does not appear in the specification, it is a claimed mutant sequence of mature cholera toxin B subunit (see AAW06605).

Sequence 103 AA:

Query Match 100.0%; Score 106; DB 17; Length 103;
Best Local Similarity 100.0%; Pred. No. 3.2e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSOHDSOKKAI 21
Db 45 gatfvevpgsqhdsqkai 65

RESULT 6
ID AAW80808 standard; protein; 103 AA.

AAW80808;
29-JAN-1999 (first entry)

Amino acid sequence of the wild type cholera toxin B subunit.

Cholera toxin B subunit; nontoxic subunit; adjuvant; coadministration;
antigen; bird; animal; mucosal; vaccine.

Vibrio cholerae.

WO9845324-A1.

15-OCT-1998.

03-APR-1998; 98WO-US06725.

04-APR-1997; 97US-0043410.

(KIYO/) KIYONO H.
(MCGH/) MCGHEE J R.
(TAKE/) TAKEDA Y.
(UABR-) UAB RES FOUND.
(YAMA/) YAMAMOTO S.

Kiyono H, Mcghee JR, Takeda Y, Yamamoto S;

WPI: 1998-594478/50.

New mutant cholera toxin selected from a group comprising nontoxic subunits/derivatives - effective as an adjuvant when coadministered with an antigen to birds and mammals

Disclosure: Fig 1B; 43pp; English.

This is the amino acid sequence of the cholera toxin B subunit used in the method of the invention involving the use of nontoxic subunits as an effective adjuvant in coadministration of an antigen to birds and animals. In addition to the use of the toxin as an mucosal adjuvant, it also provides a vaccine comprising the toxin, an immunogenic amount of an antigen, and a pharmaceutically acceptable carrier. The toxin can be used with single/multiple vaccines, and it

enables the possibility for commercial mucosal adjuvants for use in humans, since these are more effective and safer than vaccines administered subcutaneously.

Sequence 103 AA:

Query Match 100.0%; Score 106; DB 19; Length 103;
Best Local Similarity 100.0%; Pred. No. 3.2e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSOHDSOKKAI 21
Db 45 gatfvevpgsqhdsqkai 65

RESULT 7

AAW62367
ID AAW62367 standard; protein; 103 AA.

AAW62367;

29-JUN-2001 (first entry)

V. cholera cholera toxin B (CTB) protein variant (GI 1421511).

AB5B subunit protein; mutation; recombinant; cholera toxin B; CTB;
immunogen; autoimmune disease; neuron demyelinating disease; rCTB;
infection; parasitic; immunosuppressive; antiarthritic; antirheumatic;
antidiabetic; neuroprotective; vaccine; variant.

Vibrio cholerae.

WO200127144-A2.

19-APR-2001.

05-OCT-2000; 2000WO-US27607.

08-OCT-1999; 99US-0158561.

(ACTI-) ACTIVE BIOTECH AB.

Handley HH, Haaparanta T, Ewalt KL;

WPI: 2001-281974/29.

Recombinant AB5B subunit protein comprising a mutation that alters the number residues available for chemical modification, useful for covalently linking to an immunogen or vaccine which can be used for treating autoimmune diseases -

Disclosure: Page 67-68; 78pp; English.

The invention relates to a recombinant AB5B subunit protein comprising at least one mutation, where the mutation alters the number of amino acid residues available for chemical modification as compared to a wild type AB5B subunit protein, and where the recombinant protein retains an effective target ligand binding affinity. A recombinant AB5B subunit protein such as cholera toxin B protein (CTB) can be specifically covalently linked at lysines or cysteines to an immunogen or vaccine. Recombinant mutant CTB (rCTB) can be used in the treatment of autoimmune diseases e.g. rheumatoid arthritis, encephalomyelitis (or other neuron demyelinating diseases) and diabetes. The rCTB or other B-subunits of the invention can also be used to induce tolerance to infection, e.g. parasitic infection. The present sequence represents a CTB variant protein, NCBI GenBank No. GI 1421511.

Sequence 103 AA:

Query Match 100.0%; Score 106; DB 22; Length 103;
Best Local Similarity 100.0%; Pred. No. 3.2e-10;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSHIDSOKKAI 21
 |||||
 DB 45 gatfvevpgsqhidsqkkai 65

RESULT 8

AAB62370
 ID AAB62370 standard; Protein: 103 AA.

AC AAB62370;

DT 29-JUN-2001 (first entry)

DE V. cholera cholera toxin B (CTB) protein variant (GI 1421525).

KW A5B subunit protein; mutation; recombinant; cholera toxin B; CTB;
 immunogen; autoimmune disease; neuron demyelinating disease; rCTB;
 infection; parasitic; immunosuppressive; antiarthritic; antirheumatic;
 antidiabetic; neuroprotective; vaccine; variant.

OS Vibrio cholerae.

PN WO200127144-A2.

PD 19-APR-2001.

PF .05-OCT-2000; 2000WO-US27607.

PR 08-OCT-1999; 99US-0158561.

PA (ACTI-) ACTIVE BIOTECH AB.

PI Handley HH, Haaparanta T, Ewalt KL;

XX WPI; 2001-281974/29.

PT Recombinant A5B subunit protein comprising a mutation that alters the
 number residues available for chemical modification, useful for
 covalently linking to an immunogen or vaccine which can be used for
 treating autoimmune diseases -

PS Disclosure: Page 69; 78pp; English.

CC The invention relates to a recombinant A5B subunit protein comprising at
 least one mutation, where the mutation alters the number of amino acid
 residues available for chemical modification as compared to a wild type
 CC A5B subunit protein, and where the recombinant protein retains an
 effective target ligand binding affinity. A recombinant A5B subunit
 CC protein such as cholera toxin B protein (CTB) can be specifically
 covalently linked at lysines or cysteines to an immunogen or vaccine.
 CC Recombinant mutant CTB (rCTB) can be used in the treatment of autoimmune
 diseases e.g. rheumatoid arthritis, encephalomyelitis (or other neuron
 demyelinating diseases) and diabetes. The rCTB or other B subunits of the
 CC invention can also be used to induce tolerance to infection, e.g.
 CC parasitic infection. The present sequence represents a CTB variant
 CC protein, NCBI Genbank No. GI 1421525 (Ogawa 41 R35D).

XX Sequence 103 AA;

Query Match 100.0%; Score 106; DB 22; Length 103;
 Best Local Similarity 100.0%; Pred. No. 3.2e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSHIDSOKKAI 21
 |||||
 DB 45 gatfvevpgsqhidsqkkai 65

RESULT 9

AAB62369

ID AAB62369 standard; Protein: 105 AA.

AC AAB62369;

DT 29-JUN-2001 (first entry)

DE V. cholera cholera toxin B (CTB) protein variant (GI 2781121).

KW A5B subunit protein; mutation; recombinant; cholera toxin B; CTB;
 immunogen; autoimmune disease; neuron demyelinating disease; rCTB;
 infection; parasitic; immunosuppressive; antiarthritic; antirheumatic;
 antidiabetic; neuroprotective; vaccine; variant.

OS Vibrio cholerae.

PN WO200127144-A2.

PD 19-APR-2001.

PF .05-OCT-2000; 2000WO-US27607.

PR 08-OCT-1999; 99US-0158561.

PA (ACTI-) ACTIVE BIOTECH AB.

PI Handley HH, Haaparanta T, Ewalt KL;

XX WPI; 2001-281974/29.

PT Recombinant A5B subunit protein comprising a mutation that alters the
 number residues available for chemical modification, useful for
 covalently linking to an immunogen or vaccine which can be used for
 treating autoimmune diseases -

PS Disclosure: Page 68-69; 78pp; English.

CC The invention relates to a recombinant A5B subunit protein comprising at
 least one mutation, where the mutation alters the number of amino acid
 residues available for chemical modification as compared to a wild type
 CC A5B subunit protein, and where the recombinant protein retains an
 effective target ligand binding affinity. A recombinant A5B subunit
 CC protein such as cholera toxin B protein (CTB) can be specifically
 covalently linked at lysines or cysteines to an immunogen or vaccine.
 CC Recombinant mutant CTB (rCTB) can be used in the treatment of autoimmune
 diseases e.g. rheumatoid arthritis, encephalomyelitis (or other neuron
 demyelinating diseases) and diabetes. The rCTB or other B subunits of the
 CC invention can also be used to induce tolerance to infection, e.g.
 CC parasitic infection. The present sequence represents a CTB variant
 CC protein, NCBI Genbank No. GI 2781121 (Ogawa 41).

XX Sequence 105 AA;

Query Match 100.0%; Score 106; DB 22; Length 105;
 Best Local Similarity 100.0%; Pred. No. 3.3e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSHIDSOKKAI 21
 |||||
 DB 46 gatfvevpgsqhidsqkkai 66

RESULT 10

AAR04163
 ID AAR04163 standard; protein: 118 AA.

AC AAR04163;

DT 10-SEP-1990 (first entry)

DE Cholera Toxin B-subunit.

XX Cholera toxin B-subunit; hybrid protein; heterologous IGA active antigen.

XX synthetic.
OS
XX
FH Key Location/Qualifiers
FT misc_difference 18..18
FT /*label= His or Tyr
FT region 1..11
FT /*label= signal peptide
FT /*note= absent from mature protein
XX
XX WO3003437-A.
XX
XX PD 05-APR-1990.
XX
XX PF 27-SEP-1989; 89WO-0000495.
XX
XX PR 27-SEP-1988; 88FR-0012627.
XX
XX PA (UYLI-) L'UNIVERSITE DE L'ETAT A LIEGE.
XX
XX PI L'Hoir C, Renard A, Martial J;
XX
XX DR WPI: 1990-132273/17.
XX N-NSDB: Q04046.
XX
XX PT New hybrid protein, useful in vaccines -
XX PT contains cholera toxin b subunit and heterologous Iga active
XX PT antigenic sequence.
XX
XX PS Disclosure: ; pp: French.
XX
XX CC Mature cholera toxin B-subunit is obtained when the signal peptide is
XX CC cleaved off. There is an Ochre codon at position 343-5; the sequence
XX CC downstream from it is part of a plasmid.
XX
XX SQ Sequence 118 AA;

Query Match 100.0%; Score 106; DB 11; Length 118;
Best Local Similarity 100.0%; Pred. No. 3.8e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSGHIDSOKKAI 21
|||||
DB 56 gatifvewpsqhdsgkai 76

RESULT 11

AAB62374
ID AAB62374 standard; Protein: 123 AA.

XX
AC AAB62374;

XX
DT 29-JUN-2001 (first entry)

XX
DE E. coli LTB protein variant (GI: 1395122).

XX
KW ABB subunit protein; mutation; recombinant; cholera toxin B; CTB;
KW immunogen; autoimmune disease; neuron demyelinating disease; rCTB; LTB;
KW infection; parastitic; immunosuppressive; antiarthritic; antirheumatic;
KW antidiabetic; neuroprotective; vaccine; heat labile toxin B; variant.
XX
OS Escherichia coli.

XX
PN WO200127144-A2.

XX
PD 19-APR-2001.

XX
PF 05-OCT-2000; 2000WO-US27607.

XX
PR 08-OCT-1999; 99US-0158561.

XX
PA (ACTI-) ACTIVE BIOTEC AB.

XX Handley HH, Haaparanta T, Ewalt KL;
XX
XX WPI: 2001-281974/29.
XX
XX PT Recombinant ABB subunit protein comprising a mutation that alters the
XX PT number residues available for chemical modification, useful for
XX PT covalently linking to an immunogen or vaccine which can be used for
XX PT treating autoimmune diseases -
XX
XX PS Disclosure; page 70; 78pp; English.
XX
XX CC The invention relates to a recombinant ABB subunit protein comprising at
XX CC least one mutation, where the mutation alters the number of amino acid
XX CC residues available for chemical modification as compared to a wild type
XX CC ABB subunit protein, and where the recombinant protein retains an
XX CC effective target ligand binding affinity. A recombinant ABB subunit
XX CC protein such as cholera toxin B protein (CTB) can be specifically
XX CC covalently linked at lysines or cysteines to an immunogen or vaccine.
XX CC Recombinant mutant CTB (rCTB) can be used in the treatment of autoimmune
XX CC diseases e.g. rheumatoid arthritis, encephalomyelitis (or other neuron
XX CC demyelinating diseases) and diabetes. The rCTB or other B subunits of the
XX CC invention can also be used to induce tolerance to infection, e.g.
XX CC parasitic infection. The present sequence represents a variant of the
XX CC E. coli heat labile toxin B (LTB) protein.
XX
XX SQ Sequence 123 AA;

Query Match 100.0%; Score 106; DB 22; Length 123;
Best Local Similarity 100.0%; Pred. No. 4e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSGHIDSOKKAI 21
|||||
DB 66 gatifvewpsqhdsgkai 86

RESULT 12
AAP93561
ID AAP93561 standard; protein: 124 AA.

XX
AC AAP93561;

XX
DT 06-JUN-1990 (first entry)

XX
DE B subunit of the heat-labile enterotoxin (LT-B) derived from E. coli.

XX
KW B subunit; heat-labile enterotoxin; LT-B; Escherichia coli; malaria;
KW circumsporozoite protein; fusion protein; live recombinant vaccine;
KW Salmonella; epitope.
XX
OS Escherichia coli.

XX
PN WO8902924-A.

XX
PD 06-APR-1989.

XX
PF 30-SEP-1988; 88WO-US03376.

XX
PR 02-OCT-1987; 87US-0104735.

XX
PA (PRAK-) PRAXIS BIOLOGICS IN.

XX
PI Brey RN, Majarian WR, Pillai S, Hockmeyer WT;

XX
WPI: 1989-114399/15.

DR N-PSDB; AAN90747.
 XX
 PT Live recombinant vaccine for malaria -
 PT comprising attenuated entero-invasive bacterium contg. DNA
 PT encoding epitope of malaria parasite
 XX
 PS Fig 3; p. 3/17; 105bp; English.
 XX
 CC In the patent, the DNA encoding LT-B is expressed as part of a fusion
 CC protein with an epitope of a malaria parasite, eg Region I or Region II
 CC from a repeat region of circumsporozoite protein antigen (CS) (AAP93560)
 CC or a repeat region of circumsporozoite protein antigen (CS) (AAP93560)
 CC attenuated *Salmonella enteritidis* under the left promoter control of
 CC lambda. Such bacteria can multiply in the host without causing disease or
 CC disorder and express CS that will induce a protective immune response
 CC against malaria and can be used in vaccines. Such vaccines can be
 CC multivalent.
 CC
 XX
 SQ Sequence 124 AA;
 XX
 Query Match 100.0%; Score 106; DB 10; Length 124;
 Best Local Similarity 100.0%; Pred. No. 4e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 GATFOVEVPGSQHTDSQKKA1 21
 ||||||||||||||||||
 DB 66 gatfgevpgsqhdsqkka1 86
 XX
 RESULT 13
 AAM06605
 ID AAM06605 standard; Protein: 124 AA.
 XX
 AC AAM06605;
 XX
 DT 06-AUG-1997 (first entry)
 XX
 DE Cholera toxin B subunit, used for hybrid immunogenic toxin production.
 XX
 KW Heat labile; tolerant; stable; hybrid; cholera toxin B subunit; CTB;
 KW heat labile enterotoxin B subunit; LTB; vaccine; immunisation;
 KW immunogen; antigen; broad spectrum; cross reactive; diarrhoea;
 KW vomiting; food poisoning.
 XX
 OS Vibrio cholerae.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT /label= sig_peptide
 FT Protein 22..124
 FT /label= mat_protein
 XX
 PN WO9634893-A1.
 XX
 PD 07-NOV-1996.
 XX
 PF 02-MAY-1996; 96WO-SE00570.
 XX
 PR 05-MAY-1995; 95SE-0001682.
 XX
 PA (HOLM/) HOLMGREN J.
 PA (LEBE/) LEBENS M R.
 XX
 PI Holmgren J, Lebens MR;
 XX
 DR WPI: 1996-506108/50.
 DR N-PSDB; AAT43575.
 XX
 PT Heat-labile enterotoxin B sub-unit and cholera toxin B sub-unit
 PT hybrid protein - opt. fused to immunogenic sequence for use in
 PT vaccines against enterotoxin-induced illness
 XX

PS Disclosure; Fig 1; 32pp; English.
 XX
 CC AAM06605 is the full (including the signal peptide) length cholera
 CC toxin B subunit (CTB), this sequence is described as unpublished in
 CC the specification. The mature CTB protein was used to create hybrid
 CC mutants, in which certain amino acids (aa) of CTB were replaced with
 CC corresponding aa from heat-labile enterotoxin B subunit (LTB), see
 CC AAM06606 and AAM06607. The specific amino acid substitutions impart
 CC LTB-specific epitope characteristics to immunogenic mature CTB. The
 CC hybrid molecules have increased cross-reactivity and are suitable
 CC for a broad spectrum vaccine to protect against enterotoxigenic
 CC illness. Immunogenic proteins comprising the hybrid molecules can be
 CC used to treat, or in a vaccine to prevent, enterotoxin-induced illness,
 CC e.g. diarrhoea and vomiting, in humans and animals.
 CC
 XX
 SQ Sequence 124 AA;
 XX
 Query Match 100.0%; Score 106; DB 17; Length 124;
 Best Local Similarity 100.0%; Pred. No. 4e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 GATFOVEVPGSQHTDSQKKA1 21
 ||||||||||||||||||
 DB 66 gatfgevpgsqhdsqkka1 86
 XX
 RESULT 14
 AAM59770
 ID AAM59770 standard; Protein: 124 AA.
 XX
 AC AAM59770;
 XX
 DT 12-OCT-1998 (first entry)
 XX
 DE Amino acid sequence of E. coli LTB.
 XX
 KW Beta-subunit of heat labile enterotoxin; LTB; fusion protein; vaccine;
 KW immunogen; antigen; inhibin; fertility; follicle stimulating hormone;
 KW FSH; sperm; ova; immune response.
 XX
 OS Escherichia coli.
 XX
 PN WO9821344-A1.
 XX
 PD 22-MAY-1998.
 XX
 PF 12-NOV-1997; 97WO-US20584.
 XX
 PR 12-NOV-1996; 96US-0747410.
 XX
 PA (UNMS) UNIV MICHIGAN STATE.
 XX
 PI Bagdasarian M, Ireland J;
 XX
 DR WPI: 1998-297947/26.
 DR N-PSDB; AAV41573.
 XX
 PT New nucleic acid encoding fusion of antigenic peptide and
 PT enterotoxin sub-unit - useful as vaccinating immunogen, particularly
 PT for increasing animal fertility by inducing antibodies against
 PT inhibin
 XX
 PS Disclosure; Fig 9; 56pp; English.
 XX
 CC This is the amino acid sequence of Escherichia coli beta-subunit of
 CC heat labile enterotoxin (LTB). It is used in the method of the
 CC invention to create fusion proteins which are useful as vaccinating
 CC immunogens. The fusion proteins are useful in vaccines, specifically
 CC where the antigenic peptide is an inhibin fragment for increasing the
 CC fertility of an animal (by increasing levels of follicle stimulating
 CC hormone (FSH) or production of sperm or ova), but more generally for
 CC inducing an immune response against the antigenic peptide. Vaccines

CC are particularly administered orally, e.g. fusion protein is expressed
 CC in edible plants or animals.
 XX
 SQ Sequence 124 AA:

Query Match 100.0%; Score 106; DB 19; Length 124;
 Best Local Similarity 100.0%; Pred. No. 4e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHDSQKKAI 21
 |||
 DB 66 gatfqvevpgsqhdsqkkai 86

RESULT 15
 AAY96652
 ID AAY96652 standard: Protein; 124 AA.
 XX
 AC AAY96652;

DT 26-SEP-2000 (first entry)

DE Plant-optimized E. coli heat labile toxin B subunit.

XX Heat-labile toxin; LT-A; LT-B; mutant; transgenic plant; vaccine; oral;
 KW adjuvant; anti-bacterial.

XX Escherichia coli.

OS Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..21

FT Protein /label= signal_peptide

FT /label= mature_protein

PN WO200037609-A2.

PD 29-JUN-2000.

PF 22-DEC-1999; 99WO-US30747.

PR 22-DEC-1998; 98US-0113507.

PA (BOYC-) BOYCE THOMPSON INST PLANT RES.

PA (MASO/) MASON H S.

PA (ARNT/) ARNTZEN C J.

PI Mason HS, Arntzen CJ;

DR WPI: 2000-442653/38.

DR N-PSDB: AAA51156.

PT New polynucleotides encoding LT-A or CT-A polypeptides for the
 PT transformation of plant cells; useful in immunogenic compositions to
 PT elicit immune responses in animals

PS Claim 4; Fig 5A-B; 103pp: English.

XX This synthetic Escherichia coli heat-labile toxin (LT) B subunit (LT-B)
 CC is encoded by a plant-codon optimized cDNA. The cDNA sequence contains
 CC plant-preferred codons and eliminates sequence motifs associated with
 CC spurious mRNA processing. The second codon is changed from AAT encoding
 CC Asn to GTG encoding Val, in order to create a NcoI restriction site at
 CC the 5' end. Novel polynucleotides encode a mutant LT-A polypeptide or a
 CC mutant Vibrio cholerae cholera toxin (CT) A subunit (CT-A) polypeptide,
 CC which have reduced enzyme activity as compared to the wild-type LT-A or
 CC CT-A polypeptide and where at least one of the codons is altered to a
 CC plant preferred codon. The polynucleotide further comprises a nucleic
 CC acid sequence encoding LT B subunit (LT-B) or a CT B subunit (CT-B). The
 CC polynucleotides are useful for the transformation of plant cells for the
 CC production of transgenic plants to produce edible vaccines, especially

CC oral vaccines in transgenic plants for the prophylactic or therapeutic
 CC treatment against E. coli or V. cholerae. The mutant polypeptides are
 CC also useful as adjuvants.
 XX
 SQ Sequence 124 AA:

Query Match 100.0%; Score 106; DB 21; Length 124;
 Best Local Similarity 100.0%; Pred. No. 4e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHDSQKKAI 21
 |||
 DB 66 gatfqvevpgsqhdsqkkai 86

Search completed: July 3, 2002, 08:54:57
 Job time: 287 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 08:55:39 ; Search time 34.88 Seconds
(without alignments)
14.706 Million cell updates/sec

Title: US-09-786-648-4

Perfect score: 106

Sequence: 1 GATFOVEYVPGSQHDSOKKAI 21

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCPTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfillseq1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106	100.0	102	3	US-08-952-337-5
2	106	100.0	102	3	US-08-952-337-6
3	106	100.0	103	2	US-08-472-171-2
4	106	100.0	103	2	US-08-894-526-2
5	106	100.0	103	2	US-09-013-047-2
6	106	100.0	103	4	US-09-374-597-2
7	106	100.0	103	4	US-09-191-852-21
8	106	100.0	103	5	PCT-US95-13376-21
9	106	100.0	123	3	US-08-952-337-1
10	106	100.0	123	3	US-08-952-337-2
11	106	100.0	124	2	US-08-747-410-2
12	101	95.3	93	2	US-08-292-968-26
13	101	95.3	93	2	US-08-467-974-26
14	101	95.3	93	2	US-08-467-536-26
15	101	95.3	93	3	US-08-467-976-26
16	101	95.3	93	4	US-09-082-514-26
17	101	95.3	371	2	US-08-878-989-2
18	94	88.7	124	1	US-08-449-045C-4
19	94	88.7	124	1	US-08-449-045C-4
20	94	88.7	124	6	5223610-3
21	51	48.1	448	2	US-08-878-989-2
22	51	48.1	448	4	US-09-272-796-2
23	51	48.1	508	4	US-09-344-700-4
24	42	39.6	853	2	US-09-027-337-2
25	41.5	39.2	1090	4	US-09-346-237-5
26	40.5	38.2	856	3	US-08-709-784-2
27	40.5	38.2	862	2	US-08-209-521-23

28	40.5	38.2	862	2	US-08-209-521-30	Sequence 30, Appl
29	40.5	38.2	862	4	US-09-059-461-2	Sequence 2, Appl
30	40.5	38.2	862	4	US-08-961-810-133	Sequence 133, App
31	40.5	38.2	862	4	US-08-352-902D-133	Sequence 133, App
32	39	36.8	251	1	US-07-956-700B-94	Sequence 94, Appl
33	39	36.8	251	1	US-08-476-537-94	Sequence 94, Appl
34	39	36.8	251	1	US-08-485-607-94	Sequence 94, Appl
35	39	36.8	251	2	US-08-475-879-94	Sequence 94, Appl
36	39	36.8	346	2	US-08-602-359A-34	Sequence 34, Appl
37	39	36.8	384	4	US-09-025-580-35	Sequence 35, Appl
38	39	36.8	427	4	US-09-025-580-36	Sequence 36, Appl
39	39	36.8	459	6	5194375-6	Patent No. 5194375
40	39	36.8	693	4	US-09-564-805-234	Sequence 234, App
41	39	36.8	775	2	US-08-714-070A-1	Sequence 1, Appl
42	38	35.8	845	1	US-08-416-950-11	Sequence 11, Appl
43	38	35.8	845	2	US-08-469-830-11	Sequence 11, Appl
44	38	35.8	1073	4	US-09-541-782-6	Sequence 6, Appl
45	38	35.8	1810	5	PCT-US95-11684-4	Sequence 4, Appl

ALIGNMENTS

```
RESULT 1
US-08-952-337-5
Sequence 5, Application US/08952337
Patent No. 6019973
GENERAL INFORMATION:
APPLICANT: Holmgren, Jan
APPLICANT: Lebens, Michael R.
TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
FILE REFERENCE: 3846/OD758
CURRENT APPLICATION NUMBER: US/08/952.337
EARLIER FILING DATE: 1998-01-05
EARLIER APPLICATION NUMBER: PCT/SE96/00570
EARLIER FILING DATE: 1996-05-02
EARLIER APPLICATION NUMBER: SE 9501682-0
EARLIER FILING DATE: 1995-05-05
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5
LENGTH: 102
TYPE: PRT
ORGANISM: Vibrio cholerae
US-08-952-337-5

Query Match 100.0%; Score 106; DB 3; Length 102;
Best Local Similarity 100.0%; Pred. No. 2.5e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEYVPGSQHDSOKKAI 21
Db 44 GATFOVEYVPGSQHDSOKKAI 64

RESULT 2
US-08-952-337-6
Sequence 6, Application US/08952337
Patent No. 6019973
GENERAL INFORMATION:
APPLICANT: Holmgren, Jan
APPLICANT: Lebens, Michael R.
TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
FILE REFERENCE: 3846/OD758
CURRENT APPLICATION NUMBER: US/08/952.337
EARLIER FILING DATE: 1998-01-05
EARLIER APPLICATION NUMBER: PCT/SE96/00570
EARLIER FILING DATE: 1996-05-02
EARLIER APPLICATION NUMBER: SE 9501682-0
EARLIER FILING DATE: 1995-05-05
```

NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 6
LENGTH: 102
TYPE: PRT
ORGANISM: Escherichia coli
US-08-952-337-6

Query Match 100.0%; Score 106; DB 3; Length 102;
Best Local Similarity 100.0%; Pred. No. 2.5e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSHIDSOKKAI 21
|||||
DB 44 GATFOVEVPGSHIDSOKKAI 64

RESULT 3
US-08-472-171-2
Sequence 2, Application US/08472171
Patent No. 5932714
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M.
APPLICANT: Yacoub, Reza K.
APPLICANT: Zealey, Gavin R.
APPLICANT: Klein, Michel H.
TITLE OF INVENTION: Expression Of Gene Products From
Genetically Manipulated Strains Of Bordetella
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 330 University Avenue, Suite 701
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,171
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/393,334
FILING DATE: 23-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-507 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: 416-595-1155
TELEFAX: 416-595-1163
TELEX: 065-24567 Simdas
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-472-171-2

Query Match 100.0%; Score 106; DB 2; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.5e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSHIDSOKKAI 21
|||||
DB 45 GATFOVEVPGSHIDSOKKAI 65

RESULT 4
US-08-894-526-2
Sequence 2, Application US/08894526
Patent No. 5942418
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M
APPLICANT: Yacoub, Reza K
APPLICANT: Zealey, Gavin R
APPLICANT: Klein, Michel H
TITLE OF INVENTION: EXPRESSION OF GENE PRODUCTS FROM
GENETICALLY MANIPULATED STRAINS OF BORDETELLA
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,526
FILING DATE: 01-DEC-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-724 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-894-526-2

Query Match 100.0%; Score 106; DB 2; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.5e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSHIDSOKKAI 21
|||||
DB 45 GATFOVEVPGSHIDSOKKAI 65

RESULT 5
US-09-013-047-2
Sequence 2, Application US/09013047
Patent No. 5998168
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M.
APPLICANT: Yacoub, Reza K.
APPLICANT: Zealey, Gavin R.
APPLICANT: Klein, Michel H.
TITLE OF INVENTION: Expression Of Gene Products From
Genetically Manipulated Strains Of Bordetella
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 330 University Avenue, 6th Floor
CITY: Toronto
STATE: Ontario
COUNTRY: Canada

ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/013,047
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/472,171
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/393,334
FILING DATE: 23-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-507 MIS:v9
TELECOMMUNICATION INFORMATION:
TELEPHONE: 416-595-1153
TELEFAX: 416-595-1153
TELEX: 065-24567 Simbas
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-013-047-2

Query Match 100.0%; Score 106; DB 2; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.5e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSQKAI 21
|||||
DB 45 GATFOVEVPGSQHIDSQKAI 65

RESULT 6
US-09-374-597-2
Sequence 2, Application US/09374597
Patent No. 6140082
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M.
APPLICANT: Yacoub, Reza K.
APPLICANT: Zealey, Gavin R.
APPLICANT: Klein, Michael H.
TITLE OF INVENTION: Expression Of Gene Products From
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 330 University Avenue, 6th Floor
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/374,597
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/393,334
FILING DATE: FEBRUARY 23, 1995

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24973
REFERENCE/DOCKET NUMBER: 1038-964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 416-595-1155
TELEFAX: 416-595-1163
TELEX: 065-24567 Simbas
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-374-597-2

Query Match 100.0%; Score 106; DB 4; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.5e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSQKAI 21
|||||
DB 45 GATFOVEVPGSQHIDSQKAI 65

RESULT 7
US-09-191-852-21
Sequence 21, Application US/09191852
Patent No. 6194560
GENERAL INFORMATION:
APPLICANT: Charles J. Arintzen, Hugh S. Mason, and Tariq A. Haq
TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski L.L.P.
STREET: 1301 McKinney, Suite 5100
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77010
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/191,852
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13376
FILING DATE: 24-OCT-1995
APPLICATION NUMBER: 08/817,906
FILING DATE: 04-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Fox, David L.
REGISTRATION NUMBER: 40,612
REFERENCE/DOCKET NUMBER: P015900S1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-651-5151
TELEFAX: 713-651-5246
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-09-191-852-21

Query Match 100.0%; Score 106; DB 4; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.5e-11;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATFOVEVPGSOHIDSOKKAI 21
|||||

Db 45 GATFOVEVPGSOHIDSOKKAI 65

RESULT 8
PCT-US95-13376-21
; Sequence 21, Application PC/TUS9513376
; GENERAL INFORMATION:
; APPLICANT: The Texas A&M University System
; APPLICANT: 310 Wisenbaker
; APPLICANT: College Station, Texas 77843-3369
; TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pravel, Hewitt, Kimball & Krieger
; STREET: 1177 West Loop South, 10th Floor
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77027-9095
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13376
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/328,716
; FILING DATE: 24-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jones, John W.
; REGISTRATION NUMBER: 31,380
; REFERENCE/DOCKET NUMBER: 36170/3P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-850-0909
; TELEFAX: 713-850-0165
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
PCT-US95-13376-21

Query Match 100.0%; Score 106; DB 5; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.5e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATFOVEVPGSOHIDSOKKAI 21
|||||

Db 45 GATFOVEVPGSOHIDSOKKAI 65

RESULT 9
US-08-952-337-1
; Sequence 1, Application US/08952337
; Patent No. 601973
; GENERAL INFORMATION:
; APPLICANT: Holmgren, Jan
; APPLICANT: Lebens, Michael R.
; TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
; TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS
; FILE REFERENCE: 3846/0D758
; CURRENT APPLICATION NUMBER: US/08/952,337
; CURRENT FILING DATE: 1998-01-05
; EARLIER APPLICATION NUMBER: PCT/SE96/00570
; EARLIER FILING DATE: 1996-05-02

; EARLIER APPLICATION NUMBER: SE 9501682-0
; EARLIER FILING DATE: 1995-05-05
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Vibrio cholerae
US-08-952-337-1

Query Match 100.0%; Score 106; DB 3; Length 123;
Best Local Similarity 100.0%; Pred. No. 3.1e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATFOVEVPGSOHIDSOKKAI 21
|||||

Db 65 GATFOVEVPGSOHIDSOKKAI 85

RESULT 10
US-08-952-337-2
; Sequence 2, Application US/08952337
; Patent No. 601973
; GENERAL INFORMATION:
; APPLICANT: Holmgren, Jan
; APPLICANT: Lebens, Michael R.
; TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
; TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS
; FILE REFERENCE: 3846/0D758
; CURRENT APPLICATION NUMBER: US/08/952,337
; CURRENT FILING DATE: 1998-01-05
; EARLIER APPLICATION NUMBER: PCT/SE96/00570
; EARLIER FILING DATE: 1996-05-02
; EARLIER APPLICATION NUMBER: SE 9501682-0
; EARLIER FILING DATE: 1995-05-05
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-952-337-2

Query Match 100.0%; Score 106; DB 3; Length 123;
Best Local Similarity 100.0%; Pred. No. 3.1e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATFOVEVPGSOHIDSOKKAI 21
|||||

Db 65 GATFOVEVPGSOHIDSOKKAI 85

RESULT 11
US-08-747-410-2
; Sequence 2, Application US/08747410
; Patent No. 5993820
; GENERAL INFORMATION:
; APPLICANT: BACDASARIAN, Michael
; APPLICANT: IRELAND, James
; TITLE OF INVENTION: CHIMERIC LTB VACCINES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 3100 No. 5993820west Center, 90 South Seventh St.
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible

```

: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ Version 1.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/747,410
: FILING DATE: 12-NOV-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Kettelberger, Denise M
: REGISTRATION NUMBER: 33,924
: REFERENCE/DOCKET NUMBER: 11526.1-US-01
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 612/371-5268
: TELEFAX: 612/332-9081
:
: TELETYPE:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 124 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: internal
: ORIGINAL SOURCE:
: US-08-747-410-2

```

Query Match 100.0%; Score 106; DB 2; Length 124;
 Best Local Similarity 100.0%; Pred. No. 3.2e-11;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GATFOVEVPGSOHIDSOKKAI 21
 ||||||||||||||||
 Db 66 GATFOVEVPGSOHIDSOKKAI 86

RESULT 12
 US-08-292-968-26
 Sequence 26 Application US/08292968
 Patent No. 5856122
 GENERAL INFORMATION:
 APPLICANT: READ, Randy J.
 APPLICANT: STEIN, Penelope E.
 APPLICANT: COCKLE, Stephen A.
 APPLICANT: COHEN, Raymond P.
 APPLICANT: LOOSMORE, Sheena
 APPLICANT: KLEIN, Michel H.
 APPLICANT: ARMSTRONG, Glen D.
 TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
 NUMBER OF SEQUENCES: 46
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sim & McBurney
 STREET: Suite 701, 330 University Avenue
 CITY: Toronto
 STATE: Ontario
 COUNTRY: Canada
 ZIP: M5G 1R7
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/292,968
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/110,947

```

: FILING DATE: 24-AUG-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/251,121
: FILING DATE: 31-MAY-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: STEWART, Michael I.
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-388
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 26:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 93 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-08-292-968-26

```

Query Match 95.3%; Score 101; DB 2; Length 93;
 Best Local Similarity 95.2%; Pred. No. 1.7e-10;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GATFOVEVPGSOHIDSOKKAI 21
 | |||||||||||||||
 Db 35 GATFOVEVPGSOHIDSOKKAI 55

RESULT 13
 US-08-467-974-26
 Sequence 26 Application US/08467974
 Patent No. 5965385
 GENERAL INFORMATION:
 APPLICANT: READ, Randy J.
 APPLICANT: STEIN, Penelope E.
 APPLICANT: COCKLE, Stephen A.
 APPLICANT: COHEN, Raymond P.
 APPLICANT: LOOSMORE, Sheena
 APPLICANT: KLEIN, Michel H.
 APPLICANT: ARMSTRONG, Glen D.
 TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
 NUMBER OF SEQUENCES: 46
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sim & McBurney
 STREET: Suite 701, 330 University Avenue
 CITY: Toronto
 STATE: Ontario
 COUNTRY: Canada
 ZIP: M5G 1R7
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/467,974
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/467,536
 FILING DATE: 06-JUN-1995
 APPLICATION NUMBER: US 08/292,968
 FILING DATE: 22-AUG-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/251,121
 FILING DATE: 31-MAY-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/110,947
 FILING DATE: 24-AUG-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: STEWART, Michael I.

REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-454 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-467-974-26

Query Match 95.3%; Score 101; DB 2; Length 93;
Best Local Similarity 95.2%; Pred. No. 1.7e-10;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATFQVEVPGSQHTDSQKAI 21
Db 35 GETFQVEVPGSQHTDSQKAI 55

RESULT 14
US-08-467-536-26
Sequence 26, Application US/08467536
Patent No. 5977304
GENERAL INFORMATION:
APPLICANT: READ, Randy J.
APPLICANT: STEIN, Penelope E.
APPLICANT: COCKLE, Stephen A.
APPLICANT: OOMEN, Raymond P.
APPLICANT: LOOSMORE, Sheena
APPLICANT: KLEIN, Michel H.
APPLICANT: ARMSTRONG, Glen D.
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
City: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,536
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/292,968
FILING DATE: 22-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/251,121
FILING DATE: 31-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/110,947
FILING DATE: 24-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-455 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-467-536-26

Query Match 95.3%; Score 101; DB 2; Length 93;
Best Local Similarity 95.2%; Pred. No. 1.7e-10;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATFQVEVPGSQHTDSQKAI 21
Db 35 GETFQVEVPGSQHTDSQKAI 55

RESULT 15
US-08-467-976-26
Sequence 26, Application US/08467976
Patent No. 6018022
GENERAL INFORMATION:
APPLICANT: READ, Randy J.
APPLICANT: STEIN, Penelope E.
APPLICANT: COCKLE, Stephen A.
APPLICANT: OOMEN, Raymond P.
APPLICANT: LOOSMORE, Sheena
APPLICANT: KLEIN, Michel H.
APPLICANT: ARMSTRONG, Glen D.
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
City: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,976
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/292,968
FILING DATE: 22-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/251,121
FILING DATE: 31-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/110,947
FILING DATE: 24-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-453 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-467-976-26

Query Match 95.3%; Score 101; DB 3; Length 93;
Best Local Similarity 95.2%; Pred. No. 1.7e-10;

Mon Jul 8 07:51:53 2002

us-09-786-648-4.ra1

Page 7

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSOKKAI 21

Db 35 GATFOVEVPGSQHIDSOKKAI 55

Search completed: July 3, 2002, 08:55:39
Job time: 214 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 09:03:20 ; Search time 401.04 Seconds
(without alignments)
18,431 Million cell updates/sec

Title: US-09-786-648-4
Perfect score: 106
Sequence: 1 GATFOVEYPSQHDSDOKKAI 21

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues
Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending Patents, AA Main:*

1: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	106	100.0	21	US-09-786-648-4	Sequence 4, Appli
2	106	100.0	103	US-08-393-334-2	Sequence 2, Appli
3	106	100.0	103	US-08-782-832-15	Sequence 15, Appli
4	106	100.0	103	US-08-817-906-21	Sequence 21, Appli
5	106	100.0	103	US-09-836-433-14	Sequence 14, Appli
6	106	100.0	116	US-09-836-433-20	Sequence 20, Appli
7	106	100.0	119	US-09-836-433-22	Sequence 22, Appli

8	106	100.0	124	1	PCT-US99-30747-55	Sequence 55, Appli
9	106	100.0	124	18	US-09-470-124-55	Sequence 55, Appli
10	106	100.0	138	13	US-08-914-479-2	Sequence 2, Appli
11	106	100.0	138	13	US-08-914-479A-2	Sequence 2, Appli
12	106	100.0	313	21	US-09-756-983-15	Sequence 15, Appli
13	106	100.0	351	21	US-09-756-983-18	Sequence 18, Appli
14	106	100.0	364	21	US-09-756-983-22	Sequence 22, Appli
15	101	95.3	21	21	US-09-786-648-5	Sequence 21, Appli
16	101	95.3	93	5	US-08-110-947-10	Sequence 10, Appli
17	101	95.3	93	5	US-08-110-947A-26	Sequence 26, Appli
18	101	95.3	93	6	US-08-251-121-26	Sequence 26, Appli
19	101	95.3	371	5	US-08-150-305A-3	Sequence 3, Appli
20	101	95.3	371	11	US-08-784-218-6	Sequence 6, Appli
21	101	95.3	371	12	US-08-829-026-5	Sequence 5, Appli
22	100	94.3	124	1	PCT-US99-30747-57	Sequence 57, Appli
23	100	94.3	124	18	US-09-470-124-57	Sequence 57, Appli
24	94	88.7	124	21	US-09-760-234-7	Sequence 7, Appli
25	94	88.7	382	1	PCT-US01-08582-3	Sequence 3, Appli
26	94	88.7	382	1	PCT-US01-08582-4	Sequence 4, Appli
27	94	88.7	461	14	US-09-051-315-2	Sequence 2, Appli
28	94	88.7	461	18	US-09-423-493-2	Sequence 2, Appli
29	94	88.7	750	18	US-09-402-100-2	Sequence 2, Appli
30	94	88.7	1338	18	US-09-402-100-4	Sequence 2, Appli
31	76	71.7	15	11	US-08-732-371-1	Sequence 1, Appli
32	76	71.7	15	11	US-08-732-371A-1	Sequence 1, Appli
33	62	58.5	12	21	US-09-786-648-3	Sequence 3, Appli
34	51	48.1	286	21	US-09-758-445-415	Sequence 415, Appli
35	51	48.1	448	21	US-09-769-970-2	Sequence 2, Appli
36	51	48.1	508	51	US-09-563-997-4	Sequence 4, Appli
37	48	45.3	484	16	US-09-252-991A-29252	Sequence 29252, A
38	47	44.3	369	21	US-09-708-427-58248	Sequence 58248, A
39	47	44.3	395	26	US-60-324-109-19516	Sequence 19516, A
40	47	44.3	433	21	US-09-708-427-58247	Sequence 58247, A
41	47	44.3	494	21	US-09-708-427-58246	Sequence 58246, A
42	46	43.4	99	21	US-09-758-463-761	Sequence 761, Appli
43	46	43.4	101	15	US-09-134-000-3773	Sequence 3773, Appli
44	46	43.4	365	20	US-09-692-058-19	Sequence 19, Appli
45	46	43.4	529	18	US-09-417-507-37800	Sequence 37800, A

ALIGNMENTS

US-09-786-648-4

RESULT 1

Sequence 4, Application US/09786648

GENERAL INFORMATION:

APPLICANT: Williams, Neil Andrew

TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vacci

FILE REFERENCE: 7438

CURRENT APPLICATION NUMBER: US/09/786,648

CURRENT FILING DATE: 2001-03-07

PRIOR APPLICATION NUMBER: PCT/GB99/02970

PRIOR FILING DATE: 1999-09-07

NUMBER OF SEQ ID NOS: 6

SOFTWARE: MS DOS

SEQ ID NO 4

LENGTH: 21

TYPE: PRT

ORGANISM: E. coli

FEATURE:

LOCATION: 45...65

OTHER INFORMATION: isolated or synthetic EXB beta4-alpha2 loop fragment derivabl

OTHER INFORMATION: human variant E. coli

US-09-786-648-4

Query Match 100.0%; Score 106; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.3e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOKKAI 21
|||||
DB 1 GATFOVEVPGSOHIDSOKKAI 21

RESULT 2

US-08-393-334-2
Sequence 2, Application US/08393334
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M.
APPLICANT: Yacoub, Reza K.
APPLICANT: Zealey, Gavin R.
APPLICANT: Klein, Michel H.
TITLE OF INVENTION: Expression Of Gene Products From
Genetically Manipulated Strains Of Bordetella
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Slim & Mcburney
STREET: 330 University Avenue, Suite 701
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/393,334
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24973
REFERENCE/DOCKET NUMBER: 1038-417
TELECOMMUNICATION INFORMATION:
TELEPHONE: 416-595-1155
TELEFAX: 416-595-1163
TELEX: 065-24567 Simbas
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-393-334-2

Query Match 100.0%; Score 106; DB 7; Length 103;
Best Local Similarity 100.0%; Pred. No. 9.8e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOKKAI 21
|||||
DB 45 GATFOVEVPGSOHIDSOKKAI 65

RESULT 3

US-08-782-832-15
Sequence 15, Application US/08782832
GENERAL INFORMATION:
APPLICANT: Arntzen, Charles J.
APPLICANT: Mason, Hugh S.
APPLICANT: Haq, Tariq A.
TITLE OF INVENTION: PRODUCTION OF AN ORALLY IMMUNOGENIC
BACTERIAL PROTEIN IN TRANSGENIC PLANTS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pravel, Hewitt, Kimball & Krieger
STREET: 1177 West Loop South, 10th floor
CITY: Houston
STATE: TX

COUNTRY: USA
ZIP: 77027-9095

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/782,832
FILING DATE: 13-JAN-1997
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/328,716
FILING DATE: 24-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Kettelberger, Denise M.
REGISTRATION NUMBER: 33,924
REFERENCE/DOCKET NUMBER: 36170/3

TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-850-0909
TELEFAX: 713-850-0165
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear

US-08-782-832-15

Query Match 100.0%; Score 106; DB 11; Length 103;
Best Local Similarity 100.0%; Pred. No. 9.8e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOKKAI 21
|||||
DB 45 GATFOVEVPGSOHIDSOKKAI 65

RESULT 4

US-08-817-906-21
Sequence 21, Application US/08817906
GENERAL INFORMATION:
APPLICANT: Charles J. Arntzen, Hugh S. Mason, John D. Clements,
APPLICANT: and Tariq A. Haq
TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski L.L.P.
STREET: 1301 McKinney, Suite 5100
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77010
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/817,906
FILING DATE: 08/04/97
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13376
FILING DATE: 24-OCT-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fox, David L.
REGISTRATION NUMBER: 40,612
REFERENCE/DOCKET NUMBER: 36170/3P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-651-5151
TELEFAX: 713-651-5246
INFORMATION FOR SEQ ID NO: 21:

SEQUENCE CHARACTERISTICS:
 ; LENGTH: 103 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 US-08-817-906-21

Query Match 100.0%; Score 106; DB 12; Length 103;
 Best Local Similarity 100.0%; Pred. No. 9.8e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSOKKAI 21
 |||
 DB 45 GATFOVEVPGSQHIDSOKKAI 65

RESULT 5
 US-09-836-433-14

; Sequence 14, Application US/09836433
 ; GENERAL INFORMATION:
 ; APPLICANT: Yuki, Yoshikazu
 ; TITLE OF INVENTION: HYBRID PROTEINS FOR AUTOIMMUNE DISEASE
 ; FILE REFERENCE: Not Assigned
 ; CURRENT APPLICATION NUMBER: US/09/836,433
 ; CURRENT FILING DATE: 2001-04-16
 ; NUMBER OF SEQ ID NOS: 49
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 14
 ; LENGTH: 103
 ; TYPE: PRT
 ; ORGANISM: Vibrio cholerae
 US-09-836-433-14

Query Match 100.0%; Score 106; DB 22; Length 103;
 Best Local Similarity 100.0%; Pred. No. 9.8e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSOKKAI 21
 |||
 DB 45 GATFOVEVPGSQHIDSOKKAI 65

RESULT 6
 US-09-836-433-20

; Sequence 20, Application US/09836433
 ; GENERAL INFORMATION:
 ; APPLICANT: Yuki, Yoshikazu
 ; TITLE OF INVENTION: HYBRID PROTEINS FOR AUTOIMMUNE DISEASE
 ; FILE REFERENCE: Not Assigned
 ; CURRENT APPLICATION NUMBER: US/09/836,433
 ; CURRENT FILING DATE: 2001-04-16
 ; NUMBER OF SEQ ID NOS: 49
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 20
 ; LENGTH: 116
 ; TYPE: PRT
 ; ORGANISM: synthetic construct
 US-09-836-433-20

Query Match 100.0%; Score 106; DB 22; Length 116;
 Best Local Similarity 100.0%; Pred. No. 1.1e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSOKKAI 21
 |||
 DB 45 GATFOVEVPGSQHIDSOKKAI 65

RESULT 7

US-09-836-433-22

; Sequence 22, Application US/09836433
 ; GENERAL INFORMATION:
 ; APPLICANT: Yuki, Yoshikazu
 ; TITLE OF INVENTION: HYBRID PROTEINS FOR AUTOIMMUNE DISEASE
 ; FILE REFERENCE: Not Assigned
 ; CURRENT APPLICATION NUMBER: US/09/836,433
 ; CURRENT FILING DATE: 2001-04-16
 ; NUMBER OF SEQ ID NOS: 49
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 22
 ; LENGTH: 119
 ; TYPE: PRT
 ; ORGANISM: synthetic construct
 US-09-836-433-22

Query Match 100.0%; Score 106; DB 22; Length 119;
 Best Local Similarity 100.0%; Pred. No. 1.2e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSOKKAI 21
 |||
 DB 45 GATFOVEVPGSQHIDSOKKAI 65

RESULT 8
 PCT-US99-30747-55

; Sequence 55, Application PCT/US9930747
 ; GENERAL INFORMATION:
 ; APPLICANT: Boyce Thompson Institute for Plant Research at COR
 ; TITLE OF INVENTION: Orally Immunogenic Bacterial Enterotoxins Expressed In
 ; FILE REFERENCE: 4868/85427
 ; CURRENT APPLICATION NUMBER: PCT/US99/30747
 ; CURRENT FILING DATE: 1999-12-22
 ; NUMBER OF SEQ ID NOS: 67
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 55
 ; LENGTH: 124
 ; TYPE: PRT
 ; ORGANISM: Escherichia coli
 PCT-US99-30747-55

Query Match 100.0%; Score 106; DB 1; Length 124;
 Best Local Similarity 100.0%; Pred. No. 1.2e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSOKKAI 21
 |||
 DB 66 GATFOVEVPGSQHIDSOKKAI 86

RESULT 9
 US-09-470-124-55

; Sequence 55, Application US/09470124
 ; GENERAL INFORMATION:
 ; APPLICANT: Arntzen
 ; TITLE OF INVENTION: Orally Immunogenic Bacterial Enterotoxins Expressed In
 ; FILE REFERENCE: 4868/84454
 ; CURRENT APPLICATION NUMBER: US/09/470,124
 ; CURRENT FILING DATE: 1999-12-24
 ; NUMBER OF SEQ ID NOS: 67
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 55
 ; LENGTH: 124
 ; TYPE: PRT
 ; ORGANISM: Escherichia coli
 US-09-470-124-55

Query Match 100.0%; Score 106; DB 18; Length 124;
Best Local Similarity 100.0%; Pred. No. 1.2e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21
|||||
Db 66 GATFOVEVPGSQHIDSOKKAI 86

RESULT 10
US-08-914-479-2
; Sequence 2, Application US/08914479
; GENERAL INFORMATION:
; APPLICANT: Albani, Salvatore
; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
; FILE REFERENCE: 48112.404C2
; CURRENT APPLICATION NUMBER: US/08/914,479
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 08/409,270
; PRIOR FILING DATE: 1995-03-23
; PRIOR APPLICATION NUMBER: 07/945,860
; PRIOR FILING DATE: 1992-09-16
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: LT-B-M24 hybrid molecule
US-08-914-479-2

Query Match 100.0%; Score 106; DB 13; Length 138;
Best Local Similarity 100.0%; Pred. No. 1.4e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21
|||||
Db 66 GATFOVEVPGSQHIDSOKKAI 86

RESULT 11
US-08-914-479A-2
; Sequence 2, Application US/08914479A
; GENERAL INFORMATION:
; APPLICANT: Albani, Salvatore
; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
; FILE REFERENCE: 48112.404C2
; CURRENT APPLICATION NUMBER: US/08/914,479A
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 08/409,270
; PRIOR FILING DATE: 1995-03-23
; PRIOR APPLICATION NUMBER: 07/945,860
; PRIOR FILING DATE: 1992-09-16
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: LT-B-M24 hybrid molecule
US-08-914-479A-2

Query Match 100.0%; Score 106; DB 13; Length 138;
Best Local Similarity 100.0%; Pred. No. 1.4e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21
|||||
Db 66 GATFOVEVPGSQHIDSOKKAI 86

RESULT 12
US-09-756-983-15
; Sequence 15, Application US/09756983
; GENERAL INFORMATION:
; APPLICANT: Albani, Salvatore
; TITLE OF INVENTION: METHODS FOR ISOLATION, QUANTIFICATION,
; FILE REFERENCE: 246/285-CIP
; CURRENT APPLICATION NUMBER: US/09/756,983
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/105,018
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 09/421,506
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: PCT/US99/2466
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion constructs with human and bacterial sequences
US-09-756-983-15

Query Match 100.0%; Score 106; DB 21; Length 313;
Best Local Similarity 100.0%; Pred. No. 4e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21
|||||
Db 255 GATFOVEVPGSQHIDSOKKAI 275

RESULT 13
US-09-756-983-18
; Sequence 18, Application US/09756983
; GENERAL INFORMATION:
; APPLICANT: Albani, Salvatore
; TITLE OF INVENTION: METHODS FOR ISOLATION, QUANTIFICATION,
; FILE REFERENCE: 246/285-CIP
; CURRENT APPLICATION NUMBER: US/09/756,983
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/105,018
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 09/421,506
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: PCT/US99/2466
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 351
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion constructs with human and bacterial sequences
US-09-756-983-18

Query Match 100.0%; Score 106; DB 21; Length 351;
Best Local Similarity 100.0%; Pred. No. 4.6e-09;

Matches	21, Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	GATFEVEVPGSQHIDSQKKAI	21					
Db	293	GATFEVEVPGSQHIDSQKKAI	313					

```

RESULT 14
US-09-756-983-22
; Sequence 22. Application US/09756983
; GENERAL INFORMATION:
; APPLICANT: Aldani, Salvatore
; TITLE OF INVENTION: METHODS FOR ISOLATION, QUANTIFICATION,
; TITLE OF INVENTION: CHARACTERIZATION AND MODULATION OF
; TITLE OF INVENTION: ANTIGEN-SPECIFIC T CELLS
; FILE REFERENCE: 246/285-CIP
; CURRENT APPLICATION NUMBER: US/09/756,983
; CURRENT FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/105,018
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 09/421,506
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: PCT/US99/2466
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion constructs with human and bacterial sequences
US-09-756-983-22

```

Query Match	100.0%	Score 106;	DB 21;	Length 364;
Best Local Similarity	100.0%	Pred. No.	4.9e-09;	
Matches 21; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

```

OY      1 GATFQVEVPGSQHIDSQKKA  21
         |||||
Db      306 GATFQVEVPGSQHIDSQKKA  326

```

```

RESULT 15
US-09-786-648-5
: Sequence 5, Application US/09786648
: GENERAL INFORMATION:
: APPLICANT: Williams, Neil Andrew
: APPLICANT: Hirst, Timothy Raymond
: TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccine
: TITLE OF INVENTION: Adjuvants
: FILE REFERENCE: 7438
: CURRENT APPLICATION NUMBER: US/09/786,648
: CURRENT FILING DATE: 2001-03-07
: PRIOR APPLICATION NUMBER: PCF/GB99/02970
: PRIOR FILING DATE: 1999-09-07
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: MS DOS
: SEQ ID NO 5
: LENGTH: 21
: TYPE: PRT
: ORGANISM: E. coli
: FEATURE:
: LOCATION: 45...65
: OTHER INFORMATION: isolated or synthetic EtXB beta4-alpha2 loop fragment derivable
: OTHER INFORMATION: porcine E. coli
US-09-786-648-5

```

Query Match	95.38;	Score 101;	DB 21;	Length 21;
Best Local Similarity	95.28;	Pred. No. 9.3e-10;		
Matches 20; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

```
QY 1 GATFOVEVPGSQHIDSQKAI 21
    | | | | | | | | | | | | |
Db 1 GETFOVEVPGSQHIDSQKAI 21
```

Search completed: July 3, 2002, 09:03:21
Job time: 621 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 09:04:12 ; Search time 46.52 Seconds

(without alignments)
45.673 Million cell updates/sec

Title: US-09-786-648-4

Sequence: 1 GATPVEVPGSQHIDSOKKAI 21

Scoring table:

BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 308740 seqs, 101176262 residues

Total number of hits satisfying chosen parameters: 308740

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Pending Patents AA.New:*
1: /cgn2_6/prodata/2/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/prodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/prodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/prodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/prodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/prodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/prodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106	100.0	103	6	US-10-110-364-10
2	106	100.0	103	6	US-10-110-364-13
3	106	100.0	105	6	US-10-110-364-12
4	106	100.0	123	6	US-10-110-364-17
5	106	100.0	124	6	US-10-110-364-2
6	106	100.0	124	6	US-10-110-364-6
7	106	100.0	124	6	US-10-110-364-11
8	106	100.0	124	6	US-10-110-364-19
9	106	100.0	124	6	US-10-110-364-21
10	106	100.0	125	6	US-10-110-364-15
11	106	100.0	138	6	US-10-141-627-2
12	101	95.3	103	6	US-10-110-364-22
13	101	95.3	123	6	US-10-110-364-15
14	101	95.3	124	6	US-10-110-364-16
15	101	95.3	124	6	US-10-110-364-18
16	101	95.3	124	6	US-10-110-364-23
17	100	94.3	103	6	US-10-110-364-7
18	97	91.5	103	6	US-10-110-364-5
19	94	88.7	124	6	US-10-110-364-4
20	94	88.7	124	6	US-10-110-364-9
21	94	88.7	124	6	US-10-110-364-3
22	94	88.7	382	5	US-09-809-033A-3
23	94	88.7	382	5	US-09-809-033A-4
24	48	45.3	747	7	US-60-360-039-7808
25	43	40.6	84	5	US-09-620-393B-3233
26	43	40.6	138	5	US-09-620-393B-3232

27	43	40.6	385	5	US-09-540-209B-5629	Sequence 5629, Ap
28	43	40.6	483	7	US-60-360-039-18750	Sequence 18750, A
29	42.5	40.1	376	7	US-60-360-039-21922	Sequence 21922, A
30	42.5	39.6	492	6	US-10-155-881-18930	Sequence 18930, A
31	42	39.6	782	1	PCT-US02-07826-312	Sequence 312, App
32	42	39.6	782	6	US-10-097-340-312	Sequence 312, App
33	42	39.6	855	1	PCT-US02-09039-2	Sequence 2, App1
34	42	39.6	855	5	US-09-654-600A-2	Sequence 2, App1
35	42	39.6	855	6	US-10-104-271-2	Sequence 2, App1
36	42	39.6	855	6	US-10-099-700A-2	Sequence 2, App1
37	42	39.6	890	7	US-60-360-039-3507	Sequence 3507, Ap
38	42	39.6	2359	6	US-10-155-881-28520	Sequence 28520, A
39	41	38.7	496	6	US-10-155-881-37499	Sequence 37499, A
40	41	38.7	549	5	US-09-540-209B-6846	Sequence 6846, Ap
41	40	37.7	192	7	US-60-360-039-10029	Sequence 10029, A
42	40	37.7	211	5	US-09-540-209B-9908	Sequence 9908, Ap
43	40	37.7	369	6	US-10-138-701-36	Sequence 36, App1
44	40	37.7	417	5	US-09-540-209B-9466	Sequence 9466, Ap
45	40	37.7	657	7	US-60-360-039-5448	Sequence 5448, Ap

ALIGNMENTS

RESULT 1
US-10-110-364-10
; Sequence 10, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110.364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)..(103)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 1421511
US-10-110-364-10

Query Match 100.0%; Score 106; DB 6; Length 103;
Best Local Similarity 100.0%; Pred. No. 5.2e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATPVEVPGSQHIDSOKKAI 21
|||||
Db 45 GATPVEVPGSQHIDSOKKAI 65

RESULT 2
US-10-110-364-13
; Sequence 13, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110.364

```

; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 103
; TYPE: PRF
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(103)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 14215235
; OTHER INFORMATION: (Ogawa 41 R35D).
US-10-110-364-13
```

```

Query Match          100.0%; Score 106; DB 6; Length 103;
Best Local Similarity 100.0%; Pred. No. 5.2e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GATFOVEVPGSOHIDSOKKAI 21
DB 45 GATFOVEVPGSOHIDSOKKAI 65
```

```

RESULT 3
US-10-110-364-12
; Sequence 12, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 105
; TYPE: PRF
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(105)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 2781121 (Ogawa
; OTHER INFORMATION: 41).
US-10-110-364-12
```

```

Query Match          100.0%; Score 106; DB 6; Length 105;
Best Local Similarity 100.0%; Pred. No. 5.3e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GATFOVEVPGSOHIDSOKKAI 21
DB 46 GATFOVEVPGSOHIDSOKKAI 66
```

```

RESULT 4
US-10-110-364-17
; Sequence 17, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
```

```

; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 123
; TYPE: PRF
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(123)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 1395122.
US-10-110-364-17
```

```

Query Match          100.0%; Score 106; DB 6; Length 123;
Best Local Similarity 100.0%; Pred. No. 6.5e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GATFOVEVPGSOHIDSOKKAI 21
DB 66 GATFOVEVPGSOHIDSOKKAI 86
```

```

RESULT 5
US-10-110-364-2
; Sequence 2, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 124
; TYPE: PRF
; ORGANISM: Vibrio cholera
US-10-110-364-2
```

```

Query Match          100.0%; Score 106; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 6.5e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GATFOVEVPGSOHIDSOKKAI 21
DB 66 GATFOVEVPGSOHIDSOKKAI 86
```

```

RESULT 6
US-10-110-364-6
; Sequence 6, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
```



```

; TITLE OF INVENTION: AB5 TOXIN B SUBUNIT MUTANTS WITH ALTERED
; TITLE OF INVENTION: CHEMICAL CONUGATION CHARACTERISTICS
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 808900.
US-10-110-364-6

```

```

Query Match          100.0%; Score 106; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 6.5e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GATFOVEVPGSQHIDSOKKAI 21
Db 66 GATFOVEVPGSQHIDSOKKAI 86

```

```

RESULT 7
US-10-110-364-11
; Sequence 11, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: AB5 TOXIN B SUBUNIT MUTANTS WITH ALTERED
; TITLE OF INVENTION: CHEMICAL CONUGATION CHARACTERISTICS
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 48890 (classic
; OTHER INFORMATION: 569B).
US-10-110-364-11

```

```

Query Match          100.0%; Score 106; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 6.5e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GATFOVEVPGSQHIDSOKKAI 21
Db 66 GATFOVEVPGSQHIDSOKKAI 86

```

```

RESULT 8
US-10-110-364-19
; Sequence 19, Application US/10110364

```

```

; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: AB5 TOXIN B SUBUNIT MUTANTS WITH ALTERED
; TITLE OF INVENTION: CHEMICAL CONUGATION CHARACTERISTICS
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 1648865
; OTHER INFORMATION: (LT87).
US-10-110-364-19

```

```

Query Match          100.0%; Score 106; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 6.5e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GATFOVEVPGSQHIDSOKKAI 21
Db 66 GATFOVEVPGSQHIDSOKKAI 86

```

```

RESULT 9
US-10-110-364-21
; Sequence 21, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: AB5 TOXIN B SUBUNIT MUTANTS WITH ALTERED
; TITLE OF INVENTION: CHEMICAL CONUGATION CHARACTERISTICS
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 408996.
US-10-110-364-21

```

```

Query Match          100.0%; Score 106; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 6.5e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GATFOVEVPGSQHIDSOKKAI 21
Db 66 GATFOVEVPGSQHIDSOKKAI 86

```

```
RESULT 10
US-10-110-364-15
; Sequence 15, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 125
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(125)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 3062900.
US-10-110-364-15
```

```
Query Match          100.0%; Score 106; DB 6; Length 125;
Best Local Similarity 100.0%; Pred. No. 6.6e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 GATFOVEVPGSOHIDSOKKAI 21
DB 67 GATFOVEVPGSOHIDSOKKAI 87

RESULT 11
US-10-141-627-2
; Sequence 2, Application US/10141627
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
; FILE REFERENCE: 481112.404C3
; CURRENT APPLICATION NUMBER: US/10/141,627
; CURRENT FILING DATE: 2002-05-07
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(123)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 223254.
US-10-141-627-2
```

```
Query Match          100.0%; Score 106; DB 6; Length 138;
Best Local Similarity 100.0%; Pred. No. 7.4e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 GATFOVEVPGSOHIDSOKKAI 21
DB 66 GATFOVEVPGSOHIDSOKKAI 86
```

```
RESULT 12
US-10-110-364-22
; Sequence 22, Application US/10110364
; GENERAL INFORMATION:
```

```
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(103)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 494265.
US-10-110-364-22
```

```
Query Match          95.3%; Score 101; DB 6; Length 103;
Best Local Similarity 95.2%; Pred. No. 3.9e-10;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 GATFOVEVPGSOHIDSOKKAI 21
DB 45 GATFOVEVPGSOHIDSOKKAI 65
```

```
RESULT 13
US-10-110-364-20
; Sequence 20, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(123)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 223254.
US-10-110-364-20
```

```
Query Match          95.3%; Score 101; DB 6; Length 123;
Best Local Similarity 95.2%; Pred. No. 4.8e-10;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 GATFOVEVPGSOHIDSOKKAI 21
DB 65 GATFOVEVPGSOHIDSOKKAI 85
```

```
RESULT 14
```

Search completed: July 3, 2002, 09:04:12
Job time: 652 sec

```

US-10-110-364-16
Sequence 16, Application US/10110364
GENERAL INFORMATION:
APPLICANT: Handley, Harold H.
APPLICANT: Haaparanta, Tapio
APPLICANT: Ewalt, Karla L.
TITLE OF INVENTION: A85 TOXIN B SUBUNIT MUTANTS WITH ALTERED
TITLE OF INVENTION: CHEMICAL COMBINATION CHARACTERISTICS
FILE REFERENCE: ACBIO.004A
CURRENT APPLICATION NUMBER: US/10/110,364
CURRENT FILING DATE: 2002-04-05
PRIOR APPLICATION NUMBER: PCT/US00/27607
PRIOR FILING DATE: 2000-10-05
PRIOR APPLICATION NUMBER: 60/158,561
PRIOR FILING DATE: 1999-10-08
NUMBER OF SEQ ID NOS: 31
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16
LENGTH: 124
TYPE: PRT
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(124)
OTHER INFORMATION: LTB variant from NCBI gene bank GI: 1169505
US-10-110-364-16

```

Query Match	95.3%	Score 101;	DB 6;	Length 124;
Best Local Similarity	95.2%	Pred. No. 4.8e-10;		
Matches 20;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

```
Db      66 GETFQVEVPGSQHIDSQKAI 86
```

```

RESULT 15
US-10-110-364-18
Sequence 18, Application US/10110364
GENERAL INFORMATION:
APPLICANT: Handley, Harold H.
APPLICANT: Haaparanta, Tapio
APPLICANT: Ewalt, Kara L.
TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS
FILE REFERENCE: ACT10.004A
CURRENT APPLICATION NUMBER: US/10/110,364
CURRENT FILING DATE: 2002-04-05
PRIOR APPLICATION NUMBER: PCT/US00/27607
PRIOR FILING DATE: 2000-10-05
PRIOR APPLICATION NUMBER: 60/158,561
PRIOR FILING DATE: 1999-10-08
NUMBER OF SEQ ID NOS: 31
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 18
LENGTH: 124
TYPE: PRT
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(124)
OTHER INFORMATION: LTB variant from NCBI gene bank GI: 145833
US-10-110-364-18

```

Query Match	95.3%	Score 101;	DB 6;	length 124;
Best Local Similarity	95.2%	Pred No. 4.8e-10;		
Matches 20; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

Db 66 GETFOVEVPGSQHIDSQKAI 86

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 08:56:32 ; Search time 46.57 Seconds
(without alignments)
43.330 Million cell updates/sec

Title: US-09-786-648-4

Perfect score: 106

Sequence: 1 GATFOVEVPGSOHDSQKKAI 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : PIR_71:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	101	95.3	124	1 QLECB	heat-labile enterotoxin B precursor - Escherichia coli
2	100	94.3	124	1 XVVCB	cholera enterotoxin
3	48	45.3	392	2 T04150	RD23 protein homo
4	45	43.4	374	2 T19866	hypothetical prote
5	46	43.4	574	2 C86400	hypothetical prote
6	44.5	42.0	374	2 D81715	conserved hypotnet
7	44.5	42.0	650	2 AG0733	probable bacterioph
8	44	41.5	91	2 C96580	hypothetical prote
9	44	41.5	255	2 A86457	probable peptide c
10	44	41.5	1742	2 S76110	hypothetical prote
11	43	40.6	365	2 H69231	sensory transducti
12	43	40.6	399	2 T46898	neutrophilic prote
13	43	40.6	439	2 T49189	kinesin heavy chai
14	43	40.6	935	2 AE2469	two-component hybr
15	42.5	40.1	376	1 S17246	chorismate synthas
16	42	39.6	216	2 F83962	hypothetical prote
17	42	39.6	299	2 JG0178	chitinase (EC 3.2.
18	42	39.6	427	2 JG5694	stress-activated p
19	42	39.6	864	2 G81692	inner membrane pro
20	42	39.6	864	2 T49574	probable carnitine
21	41.5	39.2	500	2 JG4022	4-aminobutyrate tr
22	41.5	39.2	1090	2 S11823	alpha-dextrin endo
23	41	38.7	91	2 F81109	conserved hypotnet
24	41	38.7	128	1 UQDOR	ubiquitin / riboso
25	41	38.7	154	1 UQDOR7	ubiquitin / riboso
26	41	38.7	228	2 D34080	ubiquitin 18 - sli
27	41	38.7	229	2 B27806	ubiquitin (clone 1
28	41	38.7	313	2 E88216	protein B0495.8 [i
29	41	38.7	380	2 C34080	polyubiquitin 5 (c

30	41	38.7	380	2 B34080	polyubiquitin 5 (c
31	41	38.7	381	2 A27806	polyubiquitin 5 (c
32	41	38.7	386	1 A66056	yeast protein - Bac
33	41	38.7	532	2 A34080	polyubiquitin 7 (c
34	41	38.7	550	2 T01770	hypothetical prote
35	41	38.7	2911	2 T20566	hypothetical prote
36	40.5	38.2	862	2 S47598	multi protein homol
37	40	37.7	79	2 C95152	1566 family elemen
38	40	37.7	79	2 H95167	1566 family elemen
39	40	37.7	105	2 G69903	hypothetical prote
40	40	37.7	227	2 T32894	hypothetical prote
41	40	37.7	242	2 T34767	hypothetical prote
42	40	37.7	260	2 C96827	protein F20B17.2 [
43	40	37.7	330	2 C89848	peptide chain rele
44	40	37.7	355	2 T24938	hypothetical prote
45	40	37.7	367	2 T01751	gibberellin 20-oxi

ALIGNMENTS

RESULT 1
QLECB
heat-labile enterotoxin chain B precursor - Escherichia coli
C:Species: Escherichia coli
C:Date: 29-Jun-1981 #sequence, revision 29-Jun-1981 #text, change 18-Jun-1999
C:Accession: A01820; B26946; I41194; I41287; I67644; A61475
R:Dallas, W.S.; Falkow, S.
Nature 288, 499-501, 1980
A:Title: Amino acid sequence homology between cholera toxin and Escherichia coli heat
A:Reference number: A01820; MUID:81074965
A:Accession: A01820
A:Molecule type: mRNA
A:Residues: 1-124 <DAL>
R:Yamamoto, T.; Gojobori, T.; Yokota, T.
J. Bacteriol. 169, 1352-1357, 1987
A:Title: Evolutionary origin of pathogenic determinants in enterotoxigenic Escherich
A:Reference number: A26946; MUID:87137303
A:Accession: B26946
A:Molecule type: DNA
A:Residues: 1-27, 'E', 29-63, 'K', 65-124 <YAM>
A:Cross-references: EMBL:M15363; NID:g148335; PIDN:AAA24792.1; PID:g148336
R:Leong, J.; Vinal, A.C.; Dallas, W.S.
Infect. Immun. 48, 73-77, 1985
A:Title: Nucleotide sequence comparison between heat-labile toxin B-subunit cistrons
A:Reference number: I41194; MUID:85156481
A:Accession: I41194
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5, 'E', 7-17, 'C', 19-24, 'S', 26-27, 'E', 29-33, 'H', 35-63, 'K', 65-66, 'A', 68-122
A:Cross-references: GB:M17874; NID:g145830; PIDN:AAA98064.1; PID:g145831
A:Experimental source: plasmid ENT-R PCG86
R:Ibrahim, I.; Gentz, R.
J. Biol. Chem. 262, 10189-10194, 1987
A:Title: A functional interaction between the signal peptide and the translation appa
liculum.
A:Reference number: I41287; MUID:87280041
A:Accession: I41287
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-22 <RE2>
A:Cross-references: GB:M17101; NID:g146375; PIDN:AAA23973.1; PID:g146376
R:Imoue, T.; Tsuji, T.; Koto, M.; Imanura, S.; Miyama, A.
FEBS Microbiol. Lett. 108, 157-161, 1993
A:Title: Amino acid sequence of heat-labile enterotoxin from chicken enterotoxigenic
A:Reference number: I53542; MUID:99252225
A:Accession: I53542
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-17, 'C', 19, 'Y', 21-24, 'S', 26-27, 'E', 29-63, 'K', 65-66, 'A', 68-122, 'E', 124 <R
A:Cross-references: GB:560731; NID:g408994; PIDN:AA060441.1; PID:g408996
R:Tsuji, T.; Iida, T.; Honda, T.; Miyatani, T.; Nagahama, M.; Sakurai, J.; Wada, K.;
Microb. Pathog. 2, 381-390, 1987

A:Title: A unique amino acid sequence of the B subunit of a heat-labile enterotoxin iso
A:Reference number: A61475; MUID:89180953
A:Accession: A61475
A:Molecule type: protein
A:Residues: 22-24,'S',26-27,'E',29-63,'K',65-66,'A',68-95,'A',97-122,'E',124 <ISU>
A:Experimental source: strain 240-3
C:Function: the heat-labile enterotoxin molecule contains one A chain and five or six B
A:Description: the biological activity of the toxin is produced by the A chain, which ac
C:Superfamily: cholera enterotoxin beta chain
C:Keywords: enterotoxin
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-124/Product: heat-labile enterotoxin chain B #status predicted <MAT>
F:30-107/Disulfide bonds: #status predicted

Query Match 95.3%; Score 101; DB 1; Length 124;
Best Local Similarity 95.2%; Pred. No. 2.9e-09;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSQKAI 21
Db 66 GATFOVEVPGSQHIDSQKAI 86

RESULT 2
XVVCB
Cholera enterotoxin, B chain precursor VC1456 [validated] - Vibrio cholerae (strain N169
N:Alternate names: enterotoxin beta chain
C:Species: Vibrio cholerae
C>Date: 24-Apr-1984 #sequence,revision 01-Sep-2000 #text-change 02-Feb-2001
C:Accession: S14624; S39238; S39241; H82196; JC1078; S17666; PC1010; A05130; A01819; A38
R:Dams, E.; de Wolf, M.; Dierick, W.
submitted to the EMBL Data Library, March 1991
A:Description: Correction of the cholera toxin nucleotide sequence of the Vibrio cholera
A:Reference number: S14623
A:Accession: S14624
A:Molecule type: DNA
A:Residues: 1-124 <DAM>
A:Cross-references: EMBL:X58786; NID:948420; PIDN:CAA41593.1; PID:948422
A:Experimental source: strain 2125
R:Lebens, M.; Holmgren, J.
submitted to the EMBL Data Library, November 1993
A:Description: Structure and arrangement of the Cholera toxin genes in vibrio Cholera 01
A:Reference number: S39238
A:Accession: S39238
A:Molecule type: DNA
A:Residues: 1-124 <LEB>
A:Cross-references: EMBL:X76390; NID:9433856; PIDN:CAA53973.1; PID:9433857
A:Accession: S39241
A:Molecule type: DNA
A:Residues: 1-124 <LEN>
A:Cross-references: EMBL:X76391; NID:9433859; PIDN:CAA53976.1; PID:9433861
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
Chardon, D.; Esmolaeva, M.D.; Yamaheravan, J.; Bess, S.; Qin, H.; Dragoti, I.; Sellers, F
L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: H82196
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-124 <HEI>
A:Cross-references: GB:AE004224; GB:AE003852; NID:99655952; PIDN:AAF94613.1; GSPDB:GN001
A:Experimental source: serogroup O1, strain N16961, biotype El Tor
R:Shi, C.H.; Cao, C.; Zhang, J.S.; Ma, Q.J.
Chinese Biochem. J. 9, 395-399, 1993
A:Title: Nucleotide sequence analysis of the gene encoding the classical biotype V.chol
A:Reference number: JC1078
A:Accession: JC1078
A:Molecule type: DNA
A:Residues: 1-20,'O',22-31,'O',33-38,'H',40-49,'G',51-67,'T',69-124 <SHI>
A:Experimental source: classical biotype strain 569B

R:Dams, E.; de Wolf, M.; Dierick, W.
Biochim. Biophys. Acta 1090, 139-141, 1991
A:Title: Nucleotide sequence analysis of the CT operon of the Vibrio cholerae classic
A:Reference number: S17665; MUID:91355224
A:Accession: S17666
A:Molecule type: DNA
A:Residues: 1-38,'H',40-67,'T',69-124 <DA2>
A:Cross-references: EMBL:X58785; NID:948888; PIDN:CAA41591.1; PID:948890
R:Ma, Q.J.; Liu, C.X.; Xiong, L.S.; Yu, X.Q.
Sci. Sin. B Chem. Biol. Agric. Med. Earth Sci. 34, 274-280, 1991
A:Title: B subunit of cholera toxin produced in Escherichia coli.
A:Reference number: PC1010
A:Accession: PC1010
A:Molecule type: protein
A:Residues: 22-38,'H',40-41 <MAQ>
R:Mekalanos, J.J.; Swartz, D.J.; Pearson, G.D.N.; Harford, N.; Grovne, F.; de Wilde,
Nature 306, 551-557, 1983
A:Reference number: A93320; MUID:84068199
A:Accession: A05130
A:Molecule type: DNA
A:Residues: 1-32,'S',34-74,'S',76-124 <MEK>
A:Cross-references: GB:X00171; NID:948347; PIDN:CAA24996.1; PID:9758351
R:Kurosky, A.; Markel, D.E.; Peterson, J.W.
J. Biol. Chem. 252, 7257-7264, 1977
A:Title: Covalent structure of the beta chain of cholera enterotoxin.
A:Reference number: A01819; MUID:78005537
A:Accession: A01819
A:Molecule type: protein
A:Residues: 22-38,'H',40-42,'N',44-67,'T',69-90,'N',92-124 <KUR>
R:Ali, C.Y.
J. Biol. Chem. 252, 7249-7256, 1977
A:Title: Determination of the primary structure of cholera toxin B subunit.
A:Reference number: A38033; MUID:78005536
A:Accession: A38033
A:Molecule type: protein
A:Residues: 22-38,'H',40-42,'N',44-67,'T',69,'E',71-90,'N',92-124 <LAT>
A:Note: the difference at residue 70 may be due to deamidation during preparation
R:Nakashima, Y.; Napiorkowski, P.; Schafer, D.E.; Konigsberg, W.H.
FEBS Lett. 68, 275-278, 1976
A:Title: Primary structure of the B subunit of cholera enterotoxin.
A:Reference number: A38034; MUID:77026365
A:Accession: A38034
A:Molecule type: protein
A:Residues: 22-38,'H',40-67,'T',69,'E',71,'QS',74-75,'VE',78-86,'Q',88-99,'Q',101-103
R:Takao, T.; Watanabe, H.; Shimomishi, Y.
Eur. J. Biochem. 146, 503-508, 1985
A:Title: Facile identification of protein sequences by mass spectrometry.
A:Reference number: A21910; MUID:85126976
A:Accession: A21910
A:Molecule type: protein
A:Residues: 22-38,'H',40-42,'N',44-67,'T',69-90,'N',92-124 <TAK>
A:Experimental source: biotype Inaba 569B
A:Note: Asn-65 was partially deaminated to Asp.
C:Comment: The authors translated the codon TCA for residue 33 as Tyr.
C:Genetics:
A:Gene: VC1456
A:Map position: 1
A:Complex: the cholera enterotoxin molecule contains three kinds of chains; an alpha
clate noncovalently with the subunit B, an aggregate of five beta chains
C:Function: involved in binding of the toxin to cell membranes
C:Superfamily: cholera enterotoxin beta chain
C:Keywords: enterotoxin; toxin
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-124/Product: cholera enterotoxin chain B #status predicted <MAT>
F:30-107/Disulfide bonds: #status experimental

Query Match 94.3%; Score 100; DB 1; Length 124;
Best Local Similarity 95.2%; Pred. No. 4.3e-09;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSQKAI 21

A:Molecule type: DNA
A:Residues: 1-650 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD05569.1; PID:g16503066; GSPDB:GN00176
C:Genetics:
A:Gene: SRY2026

Query Match 42.0%; Score 44.5; DB 2; Length 650;
Best Local Similarity 37.5%; Pred. No. 35;
Matches 9; Conservative 5; Mismatches 7; Indels 3; Gaps 1;
QY 1 GATFOVEVPGSGHIDSOKKAI 21
DB 464 GANIPDPICGAEPPVQHAQSAKRTL 487

RESULT 8
C96580
hypothetical protein F1511.6 (imported) - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C96580
R:Theologas, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MUID:21016719
A:Accession: C96580
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-91 <STO>
A:Cross-references: GB:AE005173; NID:g4587538; PIDN:AAD35769.1; GSPDB:GN00141
C:Genetics:
A:Gene: F1511.6
A:Map position: 1

Query Match 41.5%; Score 44; DB 2; Length 91;
Best Local Similarity 47.6%; Pred. No. 5.3;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSGHIDSOKKAI 21
DB 58 GKTFFLEVKSGSETIQVKNMI 78

RESULT 9
AB6457
probable peptide chain release factor F10C21.2 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Nov-2001
C:Accession: AB6457
R:Theologas, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MUID:21016719
A:Accession: AB6457
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-255 <STO>

A:Cross-references: GB:AE005172; NID:g10645493; PIDN:AG21605.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 41.5%; Score 44; DB 2; Length 255;
Best Local Similarity 42.1%; Pred. No. 16;
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
QY 3 TFOVEVPGSGHIDSOKKAI 21
DB 102 TFRVSGPGGGRKRSASV 120

RESULT 10
S76110
hypothetical protein - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S76110
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
s.
A:Reference number: S74322; MUID:97061201
A:Accession: S76110
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1742 <KAN>
A:Cross-references: EMBL:D63999; GB:AB001339; NID:g1001396; PIDN:BA10088.1; PID:d101
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:start codon: GTG

Query Match 41.5%; Score 44; DB 2; Length 1742;
Best Local Similarity 38.1%; Pred. No. 1.2e+02;
Matches 8; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSGHIDSOKKAI 21
DB 121 GGTFSIDVPESALVADPDRIT 141

RESULT 11
H69231
sensory transduction histidine kinase - Methanobacterium thermoautotrophicum (strain
C:Species: Methanobacterium thermoautotrophicum
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: H69231
R:Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T.
; Qiu, D.; Spadafora, R.; Vitale, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani,
K.I. S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A:Reference number: A69000; MUID:98037514
A:Accession: H69231
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-365 <MTH>
A:Cross-references: GB:AE000872; GB:AE000666; NID:g2622082; PIDN:AB85482.1; PID:g262
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH985

Query Match 40.6%; Score 43; DB 2; Length 365;
Best Local Similarity 36.8%; Pred. No. 34;
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 2 ATFOVEVPGSGHIDSOKKA 20

Mon Jul 8 07:51:55 2002

us-09-786-648-4.rpr

Page 6

Search completed: July 3, 2002, 08:56:33
Job time: 233 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 09:04:42 ; Search time 21.51 seconds

(without alignments)
37.802 Million cell updates/sec

Title: US-09-786-648-4

Sequence: 1 GATGQVEVPGSGHIDSQKAI 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	106	100.0	124	1	ELBH_ECOLI
2	101	95.3	124	1	ELBP_ECOLI
3	100	94.3	124	1	CHTB_VIBCH
4	44	41.5	467	1	IRF6_HUMAN
5	44	41.5	467	1	IRF6_MOUSE
6	43	40.6	376	1	TGT_ZYMO
7	42.5	40.1	376	1	AROC_YEAST
8	42	39.6	855	1	STI4_HUMAN
9	41.5	39.2	500	1	GABT_HUMAN
10	41.5	39.2	1090	1	PULA_KLEPN
11	41	38.7	76	1	UBI0_DICDI
12	41	38.7	313	1	YP68_CAEEL
13	41	38.7	386	1	YAAN_BACSU
14	40.5	38.2	862	1	PMS2_HUMAN
15	40	37.7	288	1	HS74_CAEEL
16	40	37.7	309	1	CLT2_RAT
17	40	37.7	355	1	YRY1_CAEEL
18	40	37.7	655	1	YDH2_SCHPO
19	40	37.7	683	1	RPC_BPPHC
20	40	37.7	992	1	DLPI_RAT
21	40	37.7	2733	1	RRPH_CVMAS
22	39.5	37.3	560	1	INRI_SHEEP
23	39	36.8	112	1	Y13K_SSV1
24	39	36.8	263	1	YC43_ODOSI
25	39	36.8	456	1	SRS4_THEAC
26	39	36.8	459	1	IL7R_MOUSE
27	39	36.8	491	1	CD5_RAT
28	39	36.8	661	1	HS7C_CAEEL
29	39	36.8	664	1	CC4_CAVANL
30	39	36.8	704	1	PCCA_RAT
31	39	36.8	775	1	THI1_SCHPO
32	39	36.8	842	1	DPOL_HPEVM
33	39	36.8	1634	1	PK3B_HUMAN

34	39	36.8	2731	1	RRPB_CVMJH	P29982 murine coro
35	38.5	36.3	293	1	CHIA_CICAR	P36908 cicer ariet
36	38.5	36.3	344	1	LICH_PSESS	P25276 pseudomonas
37	38.5	36.3	529	1	TACY_LISMO	P13128 listeria mo
38	38.5	36.3	744	1	CATA_EMENTI	P55305 emericella
39	38	35.8	76	1	UBI0_NEUCR	P13117 neurospora
40	38	35.8	148	1	IF5A_AERPE	O99453 aeropyrum p
41	38	35.8	169	1	Y459_MYCGE	O99436 mycoplasma
42	38	35.8	169	1	Y459_MYCPN	P75118 mycoplasma
43	38	35.8	205	1	HAB3_KIOLA	P40914 kluyveromyc
44	38	35.8	259	1	MOB2_YEAST	P43563 saccharomyc
45	38	35.8	359	1	THD1_SOLITU	P31212 solanum tub

ALIGNMENTS

RESULT ID	1	ELBH_ECOLI	STANDARD	PRT	124 AA.
AC	P13811;				
DT	01-JAN-1990 (Rel. 13, Created)				
DT	01-JAN-1990 (Rel. 13, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	Heat-labile enterotoxin B chain precursor (LT-B, human) (LTB-B).				
GN	ELTB OR LTPB.				
OS	Escherichia coli.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.				
OX	NCBI_TaxID=562;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=ISOLATE H74-114;				
RX	MEDLINE=85156481; PubMed=3884513;				
RA	Leong J., Vinal A.C., Dallas W.S.;				
RT	"Nucleotide sequence comparison between heat-labile toxin B-subunit				
RT	clstrons from Escherichia coli of human and porcine origin.";				
RL	Infect. Immun. 48:73-77(1985).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=ISOLATE H10407;				
RX	MEDLINE=83114628; PubMed=6759877;				
RA	Yamamoto T., Tamura T.A., Yokota T., Takano T.;				
RT	"Overlapping genes in the heat-labile enterotoxin operon originating				
RT	from Escherichia coli human strain.";				
RL	Mol. Gen. Genet. 188:356-359(1982).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=ISOLATE H10407;				
RX	MEDLINE=93252225; PubMed=8486242;				
RA	Inoue T., Tsuji T., Koto M., Imamura S., Miyama A.;				
RT	"Amino acid sequence of heat-labile enterotoxin from chicken				
RT	enterotoxigenic Escherichia coli is identical to that of human strain				
RL	H 10407.";				
RL	FEMS Microbiol. Lett. 108:157-161(1993).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=ETEC LT 87;				
RA	Gerrami Y., Desperrier J.M.;				
RT	Submitted (Oct-1996) to the EMBL/GenBank/DBJ databases.				
RN	[5]				
RP	DISCUSSION OF SEQUENCE.				
RA	MEDLINE=95349400; PubMed=7623669;				
RT	Domenghini M., Pizsa M., Jobling M.G., Holmes R.K., Rappuoli R.;				
RT	"Identification of errors among database sequence entries and				
RT	comparison of correct amino acid sequences for the heat-labile				
RT	enterotoxins of Escherichia coli and Vibrio cholerae.";				
RL	Mol. Microbiol. 15:1165-1167(1995).				
RN	[6]				
RP	X-RAY CRYSTALLOGRAPHY (3.04 ANGSTROMS).				
RA	MEDLINE=99185101; PubMed=10085117;				
RT	Matkovic-Calogovic D., Loregian A., D'Acunzio M.R., Battistutta R.,				
RT	Tossi A., Palu G., Zanotti G.;				

```
RT "Crystal structure of the B subunit of Escherichia coli heat-labile
RT enterotoxin carrying peptides with anti-herpes simplex virus type 1
RT activity."
RL J. Biol. Chem. 274:8764-8769(1999).
CC -1- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
CC -1- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M17874; AAA98064.1; -
DR EMBL: J01646; AAB02982.1; -
DR EMBL: S60731; AAC60441.1; -
DR EMBL: X83966; CAA58800.1; -
DR PDB: 1LTR; 23-MAR-99.
DR InterPro: IPR001835; Enterotoxin_B.
DR Pfam: PF01376; Enterotoxin_B.1.
DR PRINTS: PR00772; ENTEROTOXIN_B.1.
DR ProDom: PD012805; Enterotoxin_B.1.
DR Enterotoxin; Signal; 3D-structure.
FT SIGNAL 1 21
FT CHAIN 22 124 HEAT-LABILE ENTEROTOXIN B CHAIN.
FT DISULFID 30 107
FT VARIANT 6 6 F -> C (IN ISOLATE H10407).
FT VARIANT 20 20 H -> Y (IN ISOLATE H10407).
FT VARIANT 34 34 H -> R (IN ISOLATE H10407).
SQ SEQUENCE 124 AA; 14027 MW; E9F7F7C7B9D3BC47 CRC64;

Query Match 100.0%; Score 106; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 2,1e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOKKAI 21
DB 66 GATFOVEVPGSOHIDSOKKAI 86

RESULT 2
ELBP_ECOLI STANDARD: PRT; 124 AA.
AC P32890; P13768; P01557;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Heat-labile enterotoxin B chain precursor (LT-B, porcine) (LTP-B).
GN ELTB OR LTPB.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ISOLATE P307;
RX MEDLINE=81074965; PubMed=7003397;
RA Dallas W.S., Falkow S.;
RT "Amino acid sequence homology between cholera toxin and Escherichia
RT coli heat-labile toxin."
RL Nature 288:499-501(1980).
RN [2]
RP REVISIONS TO 28 AND 64.
RC STRAIN-ISOLATE P307;
RX MEDLINE=85156481; PubMed=3884513;
RA Leong J., Vinal A.C., Dallas W.S.;
RT "Nucleotide sequence comparison between heat-labile toxin B-subunit
RT cistrons from Escherichia coli of human and porcine origin."
RL Infect. Immun. 48:73-77(1985).
```

```
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-ISOLATE PC86;
RX MEDLINE=87137303; PubMed=3546273;
RA Yamamoto T., Gojbori T., Yokota T.;
RT "Evolutionary origin of pathogenic determinants in enterotoxigenic
RT Escherichia coli and Vibrio cholerae O1."
RL J. Bacteriol. 169:1352-1357(1987).
RN [4]
RP SEQUENCE OF 1-22 FROM N.A.
RX MEDLINE=87280041; PubMed=3301830;
RA Ibrahim I., Gentz R.;
RT "A functional interaction between the signal peptide and the
RT translation apparatus is detected by the use of a single point
RT mutation which blocks translocation across mammalian endoplasmic
RT reticulum."
RL J. Biol. Chem. 262:10189-10194(1987).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).
RX MEDLINE=93240541; PubMed=8478941;
RA Sixma T.K., van Zanten B.A.M., Dauter Z., Hol W.G.J.;
RT "Refined structure of Escherichia coli heat-labile enterotoxin, a
RT close relative of cholera toxin."
RL J. Mol. Biol. 230:890-918(1993).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=91238966; PubMed=2034287;
RA Sixma T.K., Pronk S.E., Kalk K.H., Martna E.S., van Zanten B.A.M.,
RA Witolt B., Hol W.G.J.;
RT "Crystal structure of a cholera toxin-related heat-labile enterotoxin
RT from E. coli."
RL Nature 351:371-377(1991).
RN [7]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=95349400; PubMed=7623669;
RA Domenighini M., Pizsa M., Jobling M.G., Holmes R.K., Rappuoli R.;
RT "Identification of errors among database sequence entries and
RT comparison of correct amino acid sequences for the heat-labile
RT enterotoxins of Escherichia coli and Vibrio cholerae."
RL Mol. Microbiol. 15:1165-1167(1995).
CC -1- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
CC -1- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M17873; AAA98065.1; -
DR EMBL: M15363; AAA24792.1; -
DR EMBL: M17101; AAA23973.1; -
DR PIR: A01820; QLECB.
DR PIR: B26946; QLECEB.
DR PDB: 1LTA; 31-JAN-94.
DR PDB: 1LTB; 31-JAN-94.
DR PDB: 1LTG; 15-SEP-95.
DR PDB: 1LTI; 17-AUG-96.
DR PDB: 1LTT; 31-JAN-94.
DR PDB: 1LTV; 07-JUL-97.
DR PDB: 1LTF; 16-JUN-97.
DR PDB: 1LTS; 03-DEC-97.
DR PDB: 1LTV; 03-DEC-97.
DR PDB: 1LTL; 20-APR-95.
DR InterPro: IPR001835; Enterotoxin_B.
DR Pfam: PF01376; Enterotoxin_B.1.
DR PRINTS: PR00772; ENTEROTOXIN_B.1.
DR ProDom: PD012805; Enterotoxin_B.1.
KW Enterotoxin; Signal; 3D-structure.
```

```

FT SIGNAL 1 21
FT CHAIN 22 124 HEAT-LABILE ENTEROTOXIN B CHAIN.
FT DISULFD 30 107
FT HELIX 26 30
FT TURN 31 32
FT STRAND 36 43
FT STRAND 47 51
FT TURN 54 55
FT STRAND 58 62
FT TURN 64 65
FT STRAND 68 71
FT TURN 76 77
FT HELIX 80 98
FT TURN 99 100
FT STRAND 103 109
FT STRAND 115 123
SQ SEQUENCE 124 AA; 14133 MW; 6DB7DE58395EA70D CRC64;

Query Match 95.3%; Score 101; DB 1; Length 124;
Best Local Similarity 95.2%; Pred. No. 1.4e-09;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSKKAI 21
   | |||||
Db 66 GETFOVEVPGSOHIDSKKAI 86

RESULT 3
CHTB_VIRCH
ID CHTB_VIRCH STANDARD: PRT: 124 AA.
AC P01556; Q9JUQ2;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cholera enterotoxin, beta chain precursor.
GN CTXB OR TOXB OR VC1456.
OS Vibriol cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84061784; PubMed=6315707;
RA Lockman H., Kaper J.B.;
RT "Nucleotide sequence analysis of the A2 and B subunits of Vibrio
RT cholerae enterotoxin."
RL J. Biol. Chem. 258:13722-13726(1983).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=EL TOR 2125;
RC MEDLINE=84068199; PubMed=6646234;
RX Mekalanos J.J., Swartz D.J., Pearson G.D.N., Harford N., Groyne F.,
RA de Wilde M.;
RT "Cholera toxin genes: nucleotide sequence, deletion analysis and
RT vaccine development.";
RL Nature 306:551-557(1983).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=EL TOR 2125;
RC Dams E., de Wolf M., Dierick W.;
RL Submitted (May-1991) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=4260B / SEROTYPE O139;
RC MEDLINE=94237453; PubMed=8181723;
RX Ledens M., Holmgren J.;
RA "Structure and arrangement of the cholera toxin genes in Vibrio
RT cholerae O139."
RL FEWS Microbiol. Lett. 117:197-202(1994).
RN [5]
RP SEQUENCE FROM N.A.
RX STRAIN=1854 / O139-BENGAL;
RC Yamamoto K., Do V.G.R.F., Xu M., Iida T., Miwatani T., Albert M.J.,
```

RA Honda T.,
RL Submitted (MAY-1994) to the EMBL/GenBank/DDBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STAIN-EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
Radoson R.J., Hatt D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
Ra Ermoлева M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.P., Nieman W.C., White O.,
SA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
RN [7]
RP SEQUENCE OF 22-124.
RX MEDLINE=78005537; PubMed=903363;
RA Kurosky A., Marzel D.E., Peterson J.W.;
RT "Covalent structure of the beta chain of cholera enterotoxin."
RL J. Biol. Chem. 252:7257-7264(1977).
RN [8]
RP SEQUENCE OF 22-124.
RX MEDLINE=78005536; PubMed=903362;
RA Lai C.-Y.;
RT "Determination of the primary structure of cholera toxin B subunit."
RL J. Biol. Chem. 252:7249-7256(1977).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX MEDLINE=94272319; PubMed=8003954;
RA Merritt E.A., Sarfaty S., van den Akker F., L'Hoir C., Martial J.A.,
RA Hol W.G.J.;
RT "Crystal structure of cholera toxin B-pentamer bound to receptor GM1
RT pentaaccharide."
RL Protein Sci. 3:166-175(1994).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RX MEDLINE=95387394; PubMed=7658472;
RA Zhang R.G., Westbrook M.L., Westbrook E.M., Scott D.L., Otwinowski Z.,
RA Malik P.R., Reed R.A., Shipley G.G.;
RT "The 2.4 Å crystal structure of cholera toxin B subunit pentamer:
RT choleragenoid.";
RL J. Mol. Biol. 251:550-562(1995).
RN [11]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RC STRAIN-OGAWA 41 / CLASSICAL BIOTYPE;
RX MEDLINE=97376625; PubMed=9232653;
RA Merritt E.A., Sarfaty S., Jobling M.G., Chang T., Holmes R.K.,
RA Hirst T.R., Hol W.G.J.;
RT "Structural studies of receptor binding by cholera toxin mutants."
RL Protein Sci. 6:1516-1528(1997).
CC -I- FUNCTION: THE BETA CHAIN AGGREGATE (B SUBUNIT) IS INVOLVED IN
CC BINDING TO CELL MEMBRANES.
CC -I- SUBUNIT: CONTAINS 3 KINDS OF CHAINS. AN ALPHA AND A GAMMA CHAIN
CC (FROM THE SAME PRECURSOR MOLECULE), LINKED BY AN INTERCHAIN
CC DISULFIDE BOND, ASSOCIATE NONCOVALENTLY WITH AN AGGREGATE OF 4 TO
CC 6 BETA CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X00171; CAA2496.1; -
DR EMBL: K01170; AAA27573.1; -
DR EMBL: D30053; BAA06291.1; -
DR EMBL: X58786; CAA41593.1; -
DR EMBL: X76390; CAA53973.1; -
DR EMBL: X76391; CAA53976.1; -

```

DR EMBL: AE004224; AAF94613.1; -.
DR PIR: A01819; XVCB.
DR PIR: A05130; A05130.
DR PIR: S14624; S14624.
DR PDB: 2CHB; 03-DEC-97.
DR PDB: 3CHB; 12-AUG-98.
DR PDB: 1CHQ; 08-MAR-96.
DR PDB: 1CHQ; 08-MAR-96.
DR PDB: 1FGB; 23-DEC-96.
DR PDB: 1XPB; 01-APR-97.
DR PDB: 1XTC; 01-AUG-96.
DR PDB: 1CT1; 15-OCT-97.
DR TIGR: VC1456; -.
DR InterPro: IPR001835; Enterotoxin_B.
DR Pfam: PF01376; Enterotoxin_B; 1.
DR PRINTS: PR00772; ENTEROTOXINB.
DR ProDom: PD012805; Enterotoxin_B; 1.
KW Membrane; Enterotoxin; Signal; 3D-structure; Complete proteome.
FT SIGNAL 1 21
FT CHAIN 1 21
FT DISULFID 22 124
FT DISULFID 30 107
FT CONFLICT 33 33
FT CONFLICT 39 39
FT CONFLICT 43 43
FT CONFLICT 68 68
FT CONFLICT 70 70
FT CONFLICT 75 75
FT CONFLICT 91 91
FT HELIX 26 30
FT TURN 31 32
FT TURN 34 35
FT STRAND 36 44
FT STRAND 47 51
FT TURN 54 55
FT STRAND 58 62
FT TURN 64 65
FT STRAND 68 71
FT TURN 76 77
FT TURN 80 99
FT HELIX 100 100
FT STRAND 102 109
FT STRAND 115 123
FT STRAND 124 AA; 13957 MM; 9AA393E3EA8E3EBF CRC64;
SQ SEQUENCE 124 AA; 13957 MM; 9AA393E3EA8E3EBF CRC64;

Query Match 94.3%; Score 100; DB 1; Length 124;
Best Local Similarity 95.2%; Pred. No. 2e-09;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSHIDSOKRAI 21
DB 66 GATFOVEVPGSHIDSOKRAI 86

RESULT 4
ID IRF6_HUMAN STANDARD; PRT; 467 AA.
AC O14896;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Interferon regulatory factor 6 (IRF-6).
GN IRF6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBT_TaxID=9606;
RN NCBT_TaxID=9606;
RA SEQUENCE FROM N.A.
RA Grossman A., Miltucker H.W., Antonio L., Ozato K., Mak T.W.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.

```

```

RA Grafton D.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: BELONGS TO THE IRF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF027292; AAB84111.1; -.
DR EMBL: AL022398; CAA18545.1; -.
DR HSSP: P23906; 2IRF.
DR InterPro: IPR001346; IRF.
DR Pfam: PF00605; IRF; 1.
DR PRINTS: PR00267; INTERNREGCT.
DR ProDom: PD002355; IRF; 1.
DR SMART; SM00348; IRF; 1.
DR PROSITE; PS00601; IRF; 1.
KW Transcription regulation; DNA-binding; Nuclear protein.
FT DNA_BIND 9 111
FT SEQUENCE 467 AA; 53129 MM; 7E28F5E0F5BA4053 CRC64;
SQ SEQUENCE 467 AA; 53129 MM; 7E28F5E0F5BA4053 CRC64;

Query Match 41.5%; Score 44; DB 1; Length 467;
Best Local Similarity 41.2%; Pred. No. 12;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 5 OVEVPGSHIDSOKRAI 21
DB 273 OVEVPGSHIDSOKRAI 289

RESULT 5
ID IRF6_MOUSE STANDARD; PRT; 467 AA.
AC P97431;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Interferon regulatory factor 6 (IRF-6).
GN IRF6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBT_TaxID=10090;
RN NCBT_TaxID=10090;
RA SEQUENCE FROM N.A.
RA Grossman A., Miltucker H.W., Antonio L., Mak T.W.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: BELONGS TO THE IRF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U73029; AAB36714.1; -.
DR HSSP: P23906; 2IRF.
DR MGD; MGI:1859211; Irf6.
DR InterPro: IPR001346; IRF.
DR Pfam: PF00605; IRF; 1.
DR PRINTS: PR00267; INTERNREGCT.
DR ProDom: PD002355; IRF; 1.
DR SMART; SM00348; IRF; 1.

```

DR PROSITE: PS00601; TRF: 1.
KW Transcription regulation; DNA-binding; Nuclear protein
FT 9 111 TRYPTOPHAN PENTAD REPEAT.
FT DNA_BIND 9 111 TRYPTOPHAN PENTAD REPEAT.
SQ SEQUENCE 467 AA; 53106 MW; 68CCAA9680BEDC8 CRC64;

Query Match	41.5%	Score 44	DB 1	Length 467
Best Local Similarity	41.2%	Pred. No. 12		
Matches 7	Conservative 6	Mismatches 4	Indels 0	Gaps 0

```
OY      5 QVEVPGSQHIDSQKKA I 21
         ||: ||: ||: ||: :
Db      273 QVKFPGPEHITNEKQKL 289
```

RESULT	6
ID	TGT_ZYMO
AC	P28720; 060247; STANDARD; PRT; 385 AA.
DT	01-DEC-1992 (Rel. 24, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	01-MAR-2002 (Rel. 41, Last annotation update)
DE	Queuine tRNA-ribosyltransferase (EC 2.4.2.29) (tRNA-guanine transglycosylase) (Guanine insertion enzyme).
DE	TGT.
OS	Zymomonas mobilis.
OC	Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae; Zymomonas.
OC	NCBI_TaxID=542;
RN	[1]
RP	SEQUENCE FROM N.A., SEQUENCE OF 1-6, AND CHARACTERIZATION.
RX	MEDLINE=95394847; Pubmed=7655516;
RA	Reuter K.K.H., Flicner R.;
RT	"Sequence analysis and overexpression of the Zymomonas mobilis tgt gene encoding tRNA-guanine transglycosylase: purification and biochemical characterization of the enzyme.";
RL	J. Bacteriol. 177:5284-5288(1995).
RN	[2]
RP	PARTIAL SEQUENCE FROM N.A.
RC	STRAIN=ATCC 31821 / ZM4 / CP4;
RX	MEDLINE=92406015; Pubmed=1526462;
RA	Shark K.B., Conway T.;
RT	"Cloning and molecular characterization of the DNA ligase gene (lig) from Zymomonas mobilis.";
RL	FEMS Microbiol. Lett. 75:19-26(1992).
RN	[3]
RP	X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS).
RC	MEDLINE=96256303; Pubmed=8654383;
RA	Romler C., Reuter K., Suck D., Flicner R.;
RT	"Crystal structure of tRNA-guanine transglycosylase: RNA modification by base exchange.";
RL	EMBO J. 15:2850-2857(1996).
RN	[4]
RP	X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).
RC	MEDLINE=21103716; Pubmed=11178905;
RA	Graedler U., Gaeber H.-D., Goodenough-Dashua D.M., Garcia G.A., Flicner R., Reuter K., Stubbs M.T., Klebe G.;
RT	"A new target for shigellosis: rational design and crystallographic studies of inhibitors of tRNA-guanine transglycosylase.";
RL	J. Mol. Biol. 306:455-467(2001).
CC	-1- FUNCTION: Exchanges the guanine residue with 7-aminomethyl-7-deazaguanine in tRNAs with G(N) anticodons (tRNA-Asp, -Asn, -His and -Tyr). After this exchange, a cyclopropanediol moiety is attached to the 7-aminomethyl group of 7-deazaguanine, resulting in the hypermodified nucleoside queuine (Q) (7-(((4,5-cis-dihydroxy-2-cyclopenten-1-yl)amino)methyl)-7-deazaguanosine).
CC	-1- CATALYTIC ACTIVITY: tRNA guanine + queuine -> tRNA queuine + guanine.
CC	-1- COFACTOR: BINDS AND REQUIRES ZINC FOR ACTIVITY. ALSO REQUIRES MAGNESIUM.
CC	-1- SUBUNIT: MONOMER.
CC	-1- SIMILARITY: BELONGS TO THE QUEUINE tRNA-RIBOSYLTRANSFERASE FAMILY.
CC	-1- CAUTION: REF. 2 SEQUENCE DIFFERS FROM THAT SHOWN EXTENSIVELY DUE

```

CC      TO FRAMESHIFTS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
CC      DR      EMBL; L33777; AAA27704.1; ALT_INIT.
CC      DR      EMBL; L33777; AAA27705.1; ALT_INIT.
CC      DR      EMBL; Z11910; -: NOT_ANNOTATED_CDS.
CC      DR      PDB; 1PUD; 07-JUL-97.
CC      DR      PDB; 1WKD; 07-JUL-97.
CC      DR      PDB; 1WKE; 07-JUL-97.
CC      DR      PDB; 1WKF; 07-JUL-97.
CC      DR      PDB; 1ENU; 19-APR-00.
CC      DR      InterPro; IPR002616; Que-trna-ribstf.
CC      DR      Pfam; PF01702; YG7; 1.
CC      KW      Queuosine biosynthesis; Transferase; Glycosyltransferase;
CC      KW      tRNA processing; Zinc; Magnesium; 3D-structure.
CC      FT      INIT_MET 0
CC      FT      ACT_SITE 102 102
CC      FT      ACT_SITE 280 280
CC      FT      METAL 317 317 ZINC.
CC      FT      METAL 319 319 ZINC.
CC      FT      METAL 322 322 ZINC.
CC      FT      METAL 348 348 ZINC.
CC      SQ      SEQUENCE 385 AA; 42738 MW; F3D6FA270A0B23F3 CNC64;
CC
CC      Query Match 40.6%; Score 43; DB 1; Length 385;
CC      Best Local Similarity 38.1%; Pred. No. 14;
CC      Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;
CC
CC      QY      1 GATFOVEVPGSOHIDSOKKAI 21
CC      I I I : : I I I : : I
CC      DB      120 GVTFRSHLDGRHMLSPERSI 140
CC
CC      RESULT 7
CC      AROC_YEAST
CC      ID AROC_YEAST STANDARD: PRT; 376 AA.
CC      AC P28777;
CC      DT 01-DEC-1992 (Rel. 24, Created)
CC      DT 01-DEC-1992 (Rel. 24, Last sequence update)
CC      DT 30-MAY-2000 (Rel. 39, Last annotation update)
CC      DE Chorismate synthase (EC 4.6.1.4) (5-enolpyruvylshikimate-3-phosphate
CC      DE phospholigase).
CC      DE ARO2 OR YGL148W.
CC      OS Saccharomyces cerevisiae (Baker's yeast).
CC      OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC      OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
CC      OC NCBI_TaxID=4932;
CC      RX [1]
CC      RX SEQUENCE FROM N.A.
CC      RC STRAIN=S288C / FY1769;
CC      RC MEDLINE=97197983; PubMed=1837329;
CC      RA Jones D.G.L., Reusser U., Braus G.H.;
CC      RA "Molecular cloning, characterization and analysis of the regulation
CC      RA of the ARO2 gene, encoding chorismate synthase, of Saccharomyces
CC      RA cerevisiae.";
CC      RL Mol. Microbiol. 5:2143-2152(1991).
CC      RL [2]
CC      RP SEQUENCE FROM N.A.
CC      RC STRAIN=S288C / FY1769;
CC      RC MEDLINE=97197983; PubMed=9046099;
CC      RA Voet M., DeGoor E., Verhasselt P., Riles L., Robben J., Voelckaert G.;
CC      RA "The sequence of a nearly unclonable 22.8 kb segment on the left arm
CC      RA chromosome VII from Saccharomyces cerevisiae reveals ARO2, RPL9A,
CC      RA TIR1, MRF1 genes and six new open reading frames.";
CC      RL Yeast 13:177-182(1997).

```



```

CC      semialdehyde + L-glutamate.
CC      -1- CORACTOR: PYRIDOXAL PHOSPHATE.
CC      -1- SUBUNIT: MONOMER (PROBABLE).
CC      -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC      -1- TISSUE SPECIFICITY: LIVER > PANCREAS > BRAIN > KIDNEY > HEART >
CC      PLACENTA.
CC      -1- DISEASE: DEFECTS IN ABAAT ARE A CAUSE OF GABA-AT DEFICIENCY WHOSE
CC      PHENOTYPE INCLUDES PSYCHOMOTOR RETARDATION, HYPOTONIA,
CC      HYPERREFLEXIA, LETHARGY, REFRACTORY SEIZURES, AND EEG
CC      ABNORMALITIES.
CC      -1- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC      AMINOTRANSFERASES.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      DR      EMBL; L32961; AAA74449.1; -.
CC      DR      HSSP; P80147; 1GTX.
CC      DR      MIM; 137150; -.
CC      DR      InterPro; IPR000954; Aminotran_3.
CC      DR      Pfam; PF00202; aminotran_3; 1.
CC      DR      ProSITE; PS00600; AA_TRANSF_1; 1.
CC      KW      Transferrase; Aminotransferase; Pyridoxal phosphate;
CC      KW      Neurotransmitter degradation; Mitochondrion; Transit peptide;
CC      KW      Disease mutation.
CC      KW      CHAIN
CC      FT       1          28
CC      FT       BINDING
CC      FT       29          500
CC      FT       357          357
CC      FT       220          220
CC      FT       VARIANT
CC      FT       220
CC      FT       MITOCHONDRION
CC      FT       4-AMINOBUTYRATE AMINOTRANSFERASE.
CC      FT       PYRIDOXAL PHOSPHATE.
CC      FT       R -> K (IN GABA-AT DEFICIENCY; 25%
CC      FT       REDUCTION IN ACTIVITY).
CC      FT       /FTID=VAR_008883.
CC      FT       D -> H (IN REF. 2).
CC      FT       V -> L (IN REF. 2).
CC      FT       E -> G (IN REF. 2).
CC      FT       K -> Q (IN REF. 2).
CC      FT       W -> G (IN REF. 2).
CC      FT       S -> A (IN REF. 2).
CC      FT       G -> R (IN REF. 2).
CC      FT       C -> G (IN REF. 2).
CC      FT       L -> H (IN REF. 2).
CC      FT       SEQUENCE
CC      FT       500 AA; 56557 MW; 41199085693f80AD CRC64;

Query Match          39.2%, Score 41.5; DB 1; Length 500;
Best Local Similarity 55.6%; Pred. No. 34;
Matches 10; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

Cy      3 TFQVEVPGSGHDSQKA 20
      1:::||||:|1111
Db      17 TYRLLVPGSRHI-SQAAA 33

RESULT 10.
PULA_KLEPN
ID      PULA_KLEPN      STANDARD;      PRT;      1090 AA.
AC      P07206;
DT      01-APR-1988 (Rel. 07, Created)
DT      01-APR-1990 (Rel. 14, Last sequence update)
DT      01-MAR-2002 (Rel. 41, Last annotation update)
DE      Pullulanase precursor (EC 3.2.1.41) (Alpha-dextrin endo-1,6-alpha-
DE      glucosidase) (Pullulan 6-glucanohydrolase).
GN      PULA.
OS      Klebsiella pneumoniae.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Klebsiella.
OX      NCBI_TaxID=573;
RN      [1]
RP      SEQUENCE FROM N.A.

```


DR PIR: S00357; UODOR.
DR PIR: E34080; UODOR7.
DR PIR: A27806; A27806.
DR PIR: B27806; B27806.
DR PIR: A34080; A34080.
DR PIR: B34080; B34080.
DR PIR: C34080; C34080.
DR PIR: D34080; D34080.
DR HSP: P02248; 1UBI.
DR DictyDb: DD05001; ubqA.
DR InterPro: IPR000626; Ubiquitin.
DR Pfam: PF00240; ubiquitin.1.
DR PRINTS: PR00348; UBIOUITIN.
DR SMART: SM00213; UBO; 1.
DR PROSITE: PS00289; UBIOUITIN_1; 1.
DR PROSITE: PS00503; UBIOUITIN_2; 1.
KW Nuclear protein; Polyprotein.
FT SITE 48 48
FT BINDING 76 76
FT VARIANT 11 11
SQ SEQUENCE 76 AA; 8538 MW; 6427383968EAB84 CRC64;
NECESSARY FOR BRANCHED-CHAIN
MULTIUBIQUITIN ADDUCTS.
CONJUGATION TO ACCEPTOR PROTEINS.
K -> N (IN SOME CLONES REPEATS).

Query Match 38.7%; Score 41; DB 1; Length 76;
Best Local Similarity 42.9%; Pred. No. 5.3;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOKKAI 21
10 GKITLEVEGSDNIENKAKI 30

RESULT 12
ID YP68_CAEEL STANDARD; PRT; 313 AA.
AC 009217;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 37.0 kDa protein B0495.8 in chromosome II.
GN B0495.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Kirsten J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
-i SIMILARITY: SOME, TO YEAST YD1087C AND S.POMBE SPC16A11.13.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U21317; AAA62527.1; -
DR Wormpep: B0495.8; CE01766.
KW Hypothetical protein.
SQ SEQUENCE 313 AA; 36977 MW; 000D2327621BFED0 CRC64;

Query Match 38.7%; Score 41; DB 1; Length 313;
Best Local Similarity 46.2%; Pred. No. 25;
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;
OY 7 EVVPGSOHIDSOKK 19
:::||||:|::|

Db 11 QLMGSHVDNKK 23

RESULT 13
ID YAAN_BACSU STANDARD; PRT; 386 AA.
AC P37535;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 43.8 kDa protein in XPAC-ABRB intergenic region.
GN YAAN.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
RT subtilis chromosome containing the replication origin."
RL DNA Res. 1:1-14(1994).
RN [2]
RP SEQUENCE OF 1-191 FROM N.A.
RA Bookstein C., Edwards C.W., Hulet F.M.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D26185; BAA05262.1; -
DR EMBL: M96156; AAA22892.1; -
DR EMBL: Z99104; CAB11802.1; -
DR Subtilist; BG10090; YAAN.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 191 191 E -> M (IN REF. 2).
SQ SEQUENCE 386 AA; 43830 MW; D921F3A0F6845EEB CRC64;

Query Match 38.7%; Score 41; DB 1; Length 386;
Best Local Similarity 31.2%; Pred. No. 31;
Matches 5; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

OY 6 VEVPGSOHIDSOKKAI 21
22 IEIPGSEAVKAEKROV 37

RESULT 14
ID PMS2_HUMAN STANDARD; PRT; 862 AA.
AC P54278;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE PMS1 protein homolog 2 (DNA mismatch repair protein PMS2).
GN PMS2 OR PMSL2.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Endometrial tumor;
RX MEDLINE=94352394; PubMed=8072530;
RA Nicolaides N.C., Papadopoulos N., Liu B., Wei Y.-F., Carter K.C.,

```

RA Ruben S.M., Rosen C.A., Haseltine W.H., Fleischmann R.D.,
RA Frieser C.M., Adams M.D., Venter J.C., Dunlop M.G., Hamilton S.R.,
RA Petersen G.M., de la Chapelle A., Vogelstein B., Kinzler K.W.;
RT "Mutations of two PMS homologues in hereditary nonpolyposis colon
RT cancer.";
RL Nature 371:75-80(1994).
RN
RP [2]
RP SEQUENCE FROM N.A.
RA Bronner C.E., Baker S.M., Morrison P.T., Warren G., Smith L.G.,
RA Lescoe M.K., Kane M., Earlbino C., Lipford J., Lindblom A.,
RA Tannejaard P., Bollag R.J., Godwin A.R., Ward D.C.,
RA Nordenskiold M., Fishel R., Kolodner R.D., Liskay R.M.;
RL Submitted (Oct-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLY INVOLVED IN THE REPAIR OF MISMATCHES IN DNA.
CC -1- SUBUNIT: HETERODIMER OF PMS2 AND MLH1.
CC -1- DISEASE: ASSOCIATED WITH FAMILIAL HEREDITARY NONPOLYPOSIS COLON
CC CANCER (HNPPC) (LYNCH SYNDROME). HNPPC IS ONE OF THE MOST COMMON
CC GENETIC DISEASES IN THE WESTERN WORLD, AND ACCOUNTS FOR 15% OF ALL
CC COLON CANCERS.
CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MTL/HEXB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U13696; AAA63923.1; -.
DR EMBL: U14658; AAA50390.1; -.
DR HSSP: P23367; IBKN.
DR SWISS-2DPAGE: P54278; HUMAN.
DR MIM: 600259; -.
DR InterPro: IPR002099; DNA_mis_repair.
DR InterPro: IPR003594; HATPase_C.1.
DR Pfam: PF01119; DNA_mis_repair; 1.
DR Pfam: PF02518; HATPase_C.1.
DR PROSITE: PS00058; DNA_MISMATCH_REPAIR_1; 1.
DR KW DNA repair; Disease mutation; Anti-oncogene; Polymorphism.
FT VARIANT 20 R->Q.
FT
FT CONFLICT 470 470 /FTID=VAR_004469.
SQ SEQUENCE 862 AA; 95797 MW; B60A605222CBBCAC CRC64;

Query Match 38.2%; Score 40.5; DB 1; Length 862;
Best Local Similarity 75.0%; Pred. No. 88;
Matches 9; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

QY 10 GSQ-HIDSOKKA 20
DB 528 GSQEHVDSQEK 539

RESULT 15
HS74_CAEEL STANDARD: PRT: 288 AA.
AC P20163:
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Heat shock 70 kDa protein D (Fragment).
GN HSP-4 OR HSP70D.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peioderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89306577; PubMed=2744444;
RA Heschl M.F.P., Baillie D.L.;
RT "Identification of a heat-shock pseudogene from Caenorhabditis

```

```

RT elegans."
RL Genome 32:190-195(1989).
CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY. STRONG,
CC TO MAMMALIAN GRP78.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M28528; AAA28076.1; -.
DR HSSP: P19120; 3HSC.
DR InterPro: IPR000886; ER_target.
DR InterPro: IPR001023; HSP70.
DR Pfam: PF00012; HSP70; 1.
DR PROSITE: PS00014; ER_TARGET; 1.
DR PROSITE: PS00297; HSP70_1; PARTIAL.
DR PROSITE: PS00329; HSP70_2; PARTIAL.
DR PROSITE: PS01036; HSP70_3; PARTIAL.
DR ATP-binding; Heat shock; Endoplasmic reticulum; Multigene family.
FT NON_TER 1
FT SITE 285 288 PREVENT SECRETION FROM ER.
SQ SEQUENCE 288 AA; 31267 MW; 967F5A4A12FA67BF CRC64;

Query Match 37.7%; Score 40; DB 1; Length 288;
Best Local Similarity 35.3%; Pred. No. 33;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 3 TFOVEVPSQHIDSOK 19
DB 133 TFEIVNGILHVSADK 149

Search completed: July 3, 2002, 09:04:43
Job time: 583 sec

```

Mon Jul 8 07:51:56 2002

us-09-786-648-4.rsp

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 09:06:03 ; Search time 74.17 Seconds
(without alignments)
48.981 Million cell updates/sec

Title: US-09-786-648-4

Perfect score: 106

Sequence: 1 GATFOVEVPGSQHDSQKKAI 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_19:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.podent:*
13: sp.virus:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	106	100.0	124	2	Q57193
2	106	100.0	124	2	Q57193
3	106	100.0	124	2	Q57193
4	106	100.0	124	2	Q57193
5	106	100.0	124	2	Q57193
6	100	94.3	103	2	Q9R646
7	52	49.1	552	10	Q9LJB0
8	51	48.1	508	4	Q99987
9	49	46.2	428	10	Q9LGM2
10	48	45.3	392	10	Q40742
11	47	44.3	395	5	Q9NKM5
12	46.5	43.9	1166	3	Q9PAE0
13	46	43.4	293	5	Q49876
14	46	43.4	374	5	Q9U3M9
15	46	43.4	574	10	Q9SXC9
16	44.5	42.0	374	16	Q9PKY2

17	44	41.5	91	10	Q9SYF4	Q9SYF4 arabidopsis
18	44	41.5	255	10	Q9C815	Q9C815 arabidopsis
19	44	41.5	257	10	Q9C875	Q9C875 arabidopsis
20	44	41.5	330	11	Q9QZL7	Q9QZL7 mus musculus
21	44	41.5	467	6	Q9N136	Q9N136 ovis aries
22	44	41.5	467	11	Q91VD0	Q91VD0 mus musculus
23	44	41.5	1742	16	Q5583	Q5583 synechocyst
24	43.5	41.0	565	10	Q22511	Q22511 vitis vinif
25	43	40.6	298	10	Q43098	Q43098 psopocarpu
26	43	40.6	356	5	Q9NF90	Q9NF90 leishmania
27	43	40.6	365	17	Q27066	Q27066 methanother
28	43	40.6	399	2	Q9F517	Q9F517 zymomonas m
29	43	40.6	439	10	Q9LY67	Q9LY67 arabidopsis
30	42	39.6	216	16	Q9K922	Q9K922 bacillus ha
31	42	39.6	298	10	Q9SXM5	Q9SXM5 glycine max
32	42	39.6	299	10	Q9S7G9	Q9S7G9 glycine max
33	42	39.6	427	13	Q42099	Q42099 cyprinus ca
34	42	39.6	697	5	Q965W6	Q965W6 caenorhabdi
35	42	39.6	787	16	Q9PKR3	Q9PKR3 chlamydia m
36	42	39.6	802	10	Q947W6	Q947W6 oryza sativ
37	42	39.6	849	16	Q98LD6	Q98LD6 rhizobium 1
38	42	39.6	918	3	Q9P606	Q9P606 neurospora
39	42	39.6	1109	10	Q94HM6	Q94HM6 oryza sativ
40	42	39.6	1352	10	Q94EB4	Q94EB4 oryza sativ
41	42	39.6	2535	10	Q94BC7	Q94BC7 oryza sativ
42	41	38.7	91	16	Q9JRT4	Q9JRT4 neisseria m
43	41	38.7	246	5	Q9V906	Q9V906 drosophila
44	41	38.7	413	2	Q46961	Q46961 erwina chr
45	41	38.7	550	10	Q04636	Q04636 arabidopsis

ALIGNMENTS

RESULT 1	057193	PRELIMINARY:	PRT:	124 AA.
ID	057193			
AC	057193			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	CHOLERA TOXIN B PROTEIN (CTB) PRECURSOR (CTB).			
GN	CTXB.			
OS	Vibrio cholerae.			
OC	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.			
OX	NCBI TaxID=666;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	STRAIN-CLASSICAL STRAIN 569B:			
RX	MEDLINE=91355224; PubMed=1883840;			
RA	Dans E., De Wolf M., Dierick W.;			
RT	"Nucleotide sequence analysis of the CT operon of the Vibrio cholerae			
RL	classical strain 569B.";			
RL	Biochim. Biophys. Acta 1090:139-141(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RP	STRAIN-CLASSICAL BIOTYPE 569B:			
RC	Shi C., Cao C., Zhang J., Ma Q.;			
RA	Chin. Biochem. J. 9:395-399(1993).			
RL	[3]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN-CLASSICAL BIOTYPE 569B:			
RC	Xu L.;			
RL	Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; X58785; CAA41591.1; -.			
DR	EMBL; U25679; AAC34728.1; -.			
DR	EMBL; A00931; CAA0098.1; -.			
DR	HSSP; P01556; 2CBH.			
DR	InterPro; IPR001835; Enterotoxin_B.			
DR	Pfam; PF01376; Enterotoxin_B; 1.			
DR	PRINTS; PR00772; ENTEROTOXINB.			
DR	ProDom; PD012805; Enterotoxin_B; 1.			
KW	Signal.			

FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 124 CHOLERA TOXIN B PROTEIN (CTB).
SQ SEQUENCE 124 AA; 13919 MW; D6BF83FEF7924EA3 CRC64;

Query Match 100.0%; Score 106; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 2.6e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSHIDSOKKAI 21
DB 66 GATFOVEVPGSHIDSOKKAI 86

RESULT 2
ID O9RP15 PRELIMINARY; PRT; 124 AA.
AC O9RP15; 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, last sequence update)
DE CHOLERA ENTEROTOXIN B-SUBUNIT.
GN CTXB.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxId=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KNH002;
RA Shin H.J., Park Y.C., Kim Y.C.;
RT "Cloning and nucleotide sequence analysis of the virulence gene cassette from Vibrio cholerae KNIH002 isolated in Korea."
RL Misalimurhag Hojl 35:205-210(1999).
DR EMBL: AF175708; AAD51360.1; -
DR HSSP: P01556; 2CHB.
DR Interpro: IPR001835; Enterotoxin_B.
DR Pfam: PF01376; Enterotoxin_B.1.
DR PRINTS: PR00772; ENTEROTOXINB.
DR ProDom: PD012805; Enterotoxin_B.1.
SQ SEQUENCE 124 AA; 13905 MW; 23BF83FEF793E5B9 CRC64;

Query Match 100.0%; Score 106; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 2.6e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSHIDSOKKAI 21
DB 66 GATFOVEVPGSHIDSOKKAI 86

RESULT 3
ID O56635 PRELIMINARY; PRT; 124 AA.
AC O56635; 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
DE CHOLERA TOXIN B-SUBUNIT.
GN CTXB.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxId=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S7;
RA Yamamoto K., Do V.G., Xu M., Iida T., Miwatani T., Albert M.J.,
RA Honda T.;
RT "Comparison of cholera toxin genes (ctxAB) of non-O1 vibrio cholerae strains 854 (O139-bengal) and S7 (O37) from two outbreaks."
RL Submitted (Apr-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL: D30052; BAA06289.1; -
DR HSSP: P01556; 2CHB.

DR Interpro: IPR001835; Enterotoxin_B.
DR Pfam: PF01376; Enterotoxin_B.1.
DR PRINTS: PR00772; ENTEROTOXINB.
DR ProDom: PD012805; Enterotoxin_B.1.
KM signal.

FT SIGNAL 1 21 POTENTIAL.
SQ SEQUENCE 124 AA; 13871 MW; 3F87B2F297953179 CRC64;

Query Match 100.0%; Score 106; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 2.6e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSHIDSOKKAI 21
DB 66 GATFOVEVPGSHIDSOKKAI 86

RESULT 4
ID O93V32 PRELIMINARY; PRT; 124 AA.
AC O93V32; 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, last sequence update)
DE HEAT-LABILE ENTEROTOXIN B-SUBUNIT.
GN LTH B SUBUNIT.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxId=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1032 (ENTEROTOXIGENIC);
RA Komase K.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1032 (ENTEROTOXIGENIC);
RX MEDLINE=95091056; Pubmed=7998417;
RA Tamura S., Asanuma H., Tomita T., Komase K., Kawahara K., Danbara H.,
RA Hattori N., Watanabe K., Suzuki Y., Nagamine T., Aizawa C., Oya A.,
RA Kurata T.;
RT "Escherichia coli heat-labile enterotoxin B subunits supplemented with RT a trace amount of the holotoxin as an adjuvant for nasal influenza RT vaccine."
RL Vaccine 12:1083-1089(1994).
DR EMBL: AB011677; BAA25726.1; -
SQ SEQUENCE 124 AA; 14028 MW; 5346BD38B32354C2 CRC64;

Query Match 100.0%; Score 106; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 2.6e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSHIDSOKKAI 21
DB 66 GATFOVEVPGSHIDSOKKAI 86

RESULT 5
ID O94M01 PRELIMINARY; PRT; 124 AA.
AC O94M01; 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, last sequence update)
DE CHOLERA TOXIN B-SUBUNIT.
GN CTXB.
OS Vibrio phage CTX.
OC Viruses; ssDNA viruses; Inoviridae; Inovirus.
OX NCBI_TaxId=141904;
RN [1]

RP SEQUENCE FROM N.A.
 RA Bhattacharyya T., Nandy R.K., Nair G.B.:
 RT "The entire core region of the ctx-phi (ctx-prophage) in VCE 232, an
 RT environmental strain of *V. cholerae*.";
 RL EMBL: A0414369; to the EMBL/GenBank/DBJ databases.
 DR EMBL: A0414369; AAL09682.1;
 SQ SEQUENCE 124 AA; 13884 MW; 8D4C251591B16891 CRC64;

Query Match 100.0%; Score 106; DB 9; Length 124;
 Best Local Similarity 100.0%; Pred. No. 2.6e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GATFOVEVPGSOHIDSOKKAI 21
 66 GATFOVEVPGSOHIDSOKKAI 86

RESULT 6

ID Q9R646 PRELIMINARY; PRT; 103 AA.
 AC Q9R646;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE CHOLERA-LIKE ENTEROTOXIN B SUBUNIT.
 OS *Vibrio cholerae*.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; *Vibrio*.
 ON NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=95303036; PubMed=7783690.
 RA Nakashima K., Egueni Y., Nakasone N.;
 RT "Characterization of an enterotoxin produced by *Vibrio cholerae*
 RT O139.";
 RL Microbiol. Immunol. 39:87-94(1995).
 DR HSSP: P01556; 1XTC.
 DR InterPro: IPR001835; Enterotoxin_B.
 DR Pfam: PF01376; Enterotoxin_B.1.
 DR PRINTS: PR00772; ENTEROTOXIN_B.
 DR ProDom: PD012805; Enterotoxin_B.1.
 SQ SEQUENCE 103 AA; 11645 MW; 992A05C077F2F70E CRC64;

Query Match 94.3%; Score 100; DB 2; Length 103;
 Best Local Similarity 95.2%; Pred. No. 2.2e-09;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 GATFOVEVPGSOHIDSOKKAI 21
 45 GATFOVEVPGSOHIDSOKKAI 65

RESULT 7

ID Q9LJB0 PRELIMINARY; PRT; 552 AA.
 AC Q9LJB0;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE EMBL/CAB7996.1.
 OS *Arabidopsis thaliana* (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; *Arabidopsis*.
 ON NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-COLUMBIA;
 RX MEDLINE=20363099; PubMed=10907853;
 RA Nakamura Y.;
 RT "Structural analysis of *Arabidopsis thaliana* chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
 RT TAC and BAC clones.";
 RL DNA Res. 7:217-221(2000).
 DR EMBL: AP000606; BAB01195.1;
 SQ SEQUENCE 552 AA; 63036 MW; 740EA16CDBB2447 CRC64;

Query Match 49.1%; Score 52; DB 10; Length 552;
 Best Local Similarity 58.8%; Pred. No. 2.3;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 5 QVEVPGSOHIDSOKKAI 21
 147 QVEVPGSOHIDSOKKAI 163

RESULT 8

ID Q99987 PRELIMINARY; PRT; 508 AA.
 AC Q99987;
 DT 01-MAY-1997 (TREMblrel. 03, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE VRK2.
 GN VRK2.
 OS *Homo sapiens* (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 ON NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE=98008921; PubMed=9344656;
 RA Nezu J., Oku A., Jones M.H., Shimane M.;
 RT "Identification of two novel human putative serine/threonine kinases,
 RT VRK1 and VRK2, with structural similarity to Vaccinia virus B1R
 RT kinase.";
 RL Genomics 45:327-331(1997).
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL: AB000450; BAA19109.1;
 DR HSSP: Q06486; ICKI.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR Pfam: PF00069; pkinase.2.
 DR PROSITE: PS5011; PROTEIN_KINASE_DOM.1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST.1.
 DR ATP-binding; Serine/threonine-protein kinase; Transferase.
 KW SEQUENCE 508 AA; 58126 MW; 157FBF8F48511A4 CRC64;

Query Match 48.1%; Score 51; DB 4; Length 508;
 Best Local Similarity 50.0%; Pred. No. 3.1;
 Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Db 1 GATFOVEVPGSOHIDSOKKAI 20
 329 GATFOVEVPGSOHIDSOKKAI 348

RESULT 9

ID Q9LGM2 PRELIMINARY; PRT; 428 AA.
 AC Q9LGM2;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE EMBL: A0056822 (S20908).
 OS *Oryza sativa* (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eriarioidae; Oryzae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone: P0041E11.";
RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone: P0433F09.";
RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP002521; BAA96762.1; -;
DR EMBL: AP002539; BAB08201.1; -;
DR HSSP: Q06319; 1BUC.
DR InterPro: IPR001552; Acyl-CoA_dh.
DR Pfam: PF00441; Acyl-CoA_dh. 1.
DR Pfam: PF02770; Acyl-CoA_dh.M. 1.
DR Pfam: PF02771; Acyl-CoA_dh.N. 1.
DR PROSITE: PS00073; Acyl-CoA_DH_2; UNKNOWN.1.
SQ SEQUENCE 428 AA; 46132 MW; 8D34E3698A8E6367 CRC64;

Query Match 46.2%; Score 49; DB 10; Length 428;
Best Local Similarity 52.6%; Pred. No. 5.6;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 3 TFEVPEVPGSQHIDSQKAI 21
Db 183 TTTATKVPGGWHIDGQKRWI 201
RESULT 10
ID Q40742 PRELIMINARY; PRT; 392 AA.
AC Q40742;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE OSRAD23.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eriarioidae; Oryzae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NIPPONBARE;
RX MEDLINE=97369378; PubMed=9225866;
RA Schultz T.F., Quattrano R.S.;
RT "Characterization and expression of a rice RAD23 gene.";
RT Plant Mol. Biol. 34:557-562(1997).
DR EMBL: U63530; AAB65841.1; -;
DR HSSP: P54725; 1DVO.
DR InterPro: IPR000449; UBA.
DR InterPro: IPR000626; Ubiquitin.
DR Pfam: PF00627; UBA; 2.
DR Pfam: PF00240; ubiquitin; 1.
DR SMART: SM00165; UBA; 2.
DR SMART: SM00213; UBO; 1.
DR PROSITE: PS0053; UBQUITIN_2; 1.
SQ SEQUENCE 392 AA; 41754 MW; BDE08574CC7CACB CRC64;

Query Match 45.3%; Score 48; DB 10; Length 392;
Best Local Similarity 42.9%; Pred. No. 7.6;
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 1 GATFQVEVPGSQHIDSQKAI 21
Db 10 GSTFQIEVDSAKYKADVKRII 30

RESULT 11
ID Q9NKM5 PRELIMINARY; PRT; 395 AA.
AC Q9NKM5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 43.6 KDA PROTEIN.
GN BG:DS01514.3.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Y, CN BW SP;
RX MEDLINE=99403001; PubMed=10471707;
RA Ashburner M., Mitra S., Roote J., Lewis S.E., Blazej R., Davis T.,
RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA Palazzolo M., Reese M.G., Spredling A., Tsang G., Wan K., Whitelaw K.,
RA Celniker S., Rubin G.M.;
RT "An exploration of the sequence of a 2.9-kb region of the genome of
Drosophila melanogaster: the Adh region.";
RT Genetics 153:179-219(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Y, CN BW SP;
RA Celniker S.E., Agbayani A., Arcaine T.T., Baxter E., Blazej R.G.,
RA Butenoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA Houston K.A., Humast S.R., Kaira K., Kearney L., Kim E., Lee B.,
RA Lewis S., Li P., Lomston M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA Sethi H., Snir E., Swirskas R.R., Wan K.H., Weinbug T., Zhang R.,
RA Zieran L.L., Rubin G.M.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB003408; AAF44847.1; -;
DR Flybase: FBgn0028907; BG:DS01514.3.
KW Hypothetical protein.
SQ SEQUENCE 395 AA; 43561 MW; AEF1CC4ADD3DA73 CRC64;

Query Match 44.3%; Score 47; DB 5; Length 395;
Best Local Similarity 56.2%; Pred. No. 11;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 5 QVEVPEVPGSQHIDSQKAI 20
Db 329 RVSPGSTHIDADANA 344
RESULT 12
ID Q9PAE0 PRELIMINARY; PRT; 1166 AA.
AC Q9PAE0;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PHEROMONE-RESPONSIVE MAPKK KINASE UB4.
GN UB4.
OS Ustilago maydis (Smut fungus).
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX NCBI_TaxID=5270;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-1/2;
 RX MEDLINE=2031594; PubMed=10875339;
 RA Andrews D.L., Egan J.D., Mayorga M.E., Gold S.E.;
 RT "The *Ustilago maydis* ubc4 and ubc5 genes encode members of a MAP
 RL kinase cascade required for filamentous growth."
 DR Mol. Plant Microbe Interact. 13:781-786(2000).
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL: AF197562; AAF86841.1;
 DR HSSP: P24941; 1B38.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR000159; RA.
 DR InterPro: IPR002290; Ser_thr_kinase.
 DR Pfam: PF00069; pkinase; 1.
 DR SMART: SM00220; S_TKc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
 SQ SEQUENCE 1166 AA; 123298 MW; A89E0019AABD4172 CRC64;

Query Match 43.4%; Score 46; DB 3; Length 1166;
 Best Local Similarity 47.6%; Pred. No. 47;
 Matches 10; Conservative 4; Mismatches 6; Indels 1; Gaps 1;
 OY 2 ATFOVEVPGSOHIDSOKKAI 21
 DB 729 AKQVELPSGDSHDDQKKGM 749

RESULT 13
 ID 049876 PRELIMINARY; PRT; 293 AA.
 AC 049876;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE CLASS III CHITINASE PRECURSOR (EC 3.2.1.14).
 GN LUPAL.
 OS Lupinus albus (White lupine).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Genisteae; Lupinus.
 OX NCBI_TaxID=3870;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. ULTRA;
 RA Regalado A.P., Vidal S., Neves A., Ricardo C.P.P.,
 RA Rodrigues-Pousada C.;
 RT "Constitutive and stress-induced expression of a class III chitinase
 RL from *Lupinus albus*." to the EMBL/GenBank/DBJ databases.
 CC Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
 CC HYDROLASES).
 DR EMBL: Y16415; CAA76203.1;
 DR HSSP: P23472; 2HWV.
 DR InterPro: IPR001579; Chitinase_2.
 DR Pfam: PF00192; chitinase_2; 1.
 DR PROSITE: PS01095; CHITINASE_18; 1.
 KW Glycosidase; Hydrolase; Signal.
 FT SIGNAL 1 27
 FT CHAIN 28 293
 FT POTENTIAL.
 SQ SEQUENCE 293 AA; 31128 MW; DB1B21728F657F2F CRC64;

Query Match 43.4%; Score 46; DB 10; Length 293;
 Best Local Similarity 38.1%; Pred. No. 12;
 Matches 8; Conservative 4; Mismatches 9; Indels 0; Gaps 0;
 OY 1 GATFOVEVPGSOHIDSOKKAI 21
 DB 142 GIDFIEAGAOHYDELARAL 162

RESULT 14
 ID 0903M9 PRELIMINARY; PRT; 374 AA.
 AC 0903M9;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE C40H5.3 PROTEIN.
 GN C40H5.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA White S.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode *C. elegans*: A platform for
 RL investigating biology."
 DR Science 282:2012-2018(1998).
 DR EMBL: Z81482; CAB03954.2;
 SQ SEQUENCE 374 AA; 42198 MW; 67D202886DA7824 CRC64;

Query Match 43.4%; Score 46; DB 5; Length 374;
 Best Local Similarity 42.1%; Pred. No. 16;
 Matches 8; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
 OY 1 GATFOVEVPGSOHIDSOKK 19
 DB 230 GVEFHWFPGOPHAEQKE 248

RESULT 15
 ID 09SXC9 PRELIMINARY; PRT; 574 AA.
 AC 09SXC9;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE T17H3.2 PROTEIN (AT1G27520/T17H3_2).
 GN T17H3.2.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
 RA Lee J.M., Liu A., Li J., Gonzalez A., Liu K., Vaysberg M., Sakano H.,
 RA Chin C., Huang B., Choi E., Chou J., Altafi H., Araujo R., Brooks S.,
 RA Buehler E., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,
 RA Huizer L., Khan S., Kim C., Palm C., Rowley D., Shin P., Walker M.,
 RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
 RT "Arabidopsis thaliana chromosome 1 BAC T17H3 sequence."
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
 RA Bowser L., Carinini P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Natrusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 08:54:57 ; Search time 94.52 Seconds
(without alignments)
24.678 Million cell updates/sec

Title: US-09-786-648-5
Perfect score: 107
Sequence: 1 GEFPGYEVPGSHIDSKKAI 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*

- 1: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	21	AAV87463	E. coli heat labil
2	107	100.0	93	AAV87463	ADP-ribosylating t
3	107	100.0	93	AAV41816	Escherichia coli v
4	107	100.0	93	AAW95226	E. coli heat-labill
5	107	100.0	93	AAV68365	Heat labile toxin
6	107	100.0	93	AAV66238	E. coli verotoxin-1
7	107	100.0	103	AAV62379	E. coli LTB protei
8	107	100.0	123	AAV62377	E. coli LTB protei
9	107	100.0	124	AAV62373	E. coli LTB protei
10	107	100.0	124	AAV62375	E. coli LTB protei
11	107	100.0	124	AAV62380	E. coli LTB protei

12	107	100.0	134	22	AAV73241	Recombinant exotox
13	107	100.0	142	22	AAV73242	Recombinant exotox
14	107	100.0	155	22	AAV73243	Recombinant exotox
15	107	100.0	163	22	AAV73244	Recombinant exotox
16	107	100.0	371	20	AAV01300	Labile toxin (LT-B
17	107	100.0	371	20	AAV67443	C. jejuni flagellin
18	101	94.4	21	21	AAV87462	Cholera toxin B/en
19	101	94.4	103	17	AAV04857	Synthetic cholera
20	101	94.4	103	17	AAV94939	Heat labile entero
21	101	94.4	103	17	AAV06606	Cholera toxin B su
22	101	94.4	103	17	AAV06607	Cholera toxin B su
23	101	94.4	103	19	AAV80808	Amino acid sequenc
24	101	94.4	103	22	AAV62367	V. cholera cholera
25	101	94.4	103	22	AAV62370	V. cholera cholera
26	101	94.4	105	22	AAV62369	V. cholera cholera
27	101	94.4	118	11	AAV04163	Cholera Toxin B-su
28	101	94.4	123	22	AAV62374	E. coli LTB protei
29	101	94.4	124	10	AAV93561	B subunit of the h
30	101	94.4	124	17	AAV06605	Cholera toxin B su
31	101	94.4	124	19	AAV59770	Amino acid sequenc
32	101	94.4	124	21	AAV96652	Plant-optimized E.
33	101	94.4	124	22	AAV65992	Cholera toxin B su
34	101	94.4	124	22	AAV62359	V. cholera strain
35	101	94.4	124	22	AAV62363	V. cholera cholera
36	101	94.4	124	22	AAV62368	V. cholera cholera
37	101	94.4	124	22	AAV62376	E. coli LTB protei
38	101	94.4	124	22	AAV62378	E. coli LTB protei
39	101	94.4	125	22	AAV62372	E. coli LTB protei
40	101	94.4	131	11	AAV04825	LTB-CTB fusion pro
41	101	94.4	138	15	AAV50227	Sequence of LT-B-M
42	101	94.4	170	20	AAV94082	LTB-CTP fusion pro
43	101	94.4	380	22	AAV00507	E. coli heat-labill
44	101	94.4	382	22	AAV00506	E. coli heat-labill
45	101	94.4	405	12	AAV11272	HSV-1 antigen/heat

ALIGNMENTS

RESULT 1

AAV87463 standard; peptide: 21 AA.

AAV87463:

03-JUL-2000 (first entry)

E. coli heat labile enterotoxin B-derived peptide, SEQ ID NO:5.

Heat labile enterotoxin subunit B; EtxB;

beta-4-alpha-7 loop; GM-1 ganglioside receptor; immunomodulation;

adjuvant; immune disorder; diarrhoea.

Escherichia coli.

WO200014114-A1.

16-MAR-2000.

07-SEP-1999; 99WO-GB02970.

07-SEP-1998; 98GB-0019484.

(UYBR-) UNIV BRISTOL.

Williams NA, Hirst TR;

WPT: 2000-256943/22.

Derivatives of Escherichia coli heat labile enterotoxins useful as

immunomodulators and for treating diarrhoea and which do not bind the

glycolipid receptor GM-1 -

PS Disclosure; Page 15; 62pp; English.

XX The invention relates to peptide fragments of the *Escherichia coli* heat labile enterotoxin (Etx) and its closely related homologue, cholera toxin (Ctx) from *Vibrio cholerae* which do not bind to the ubiquitous GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are composed of one A subunit and five identical B subunits. The A subunit is responsible for toxicity, possessing adenosine diphosphate (ADP)-ribosyltransferase activity, while the B subunits (CtxB and CtxB) facilitate the entry of subunit A into the host cell via the binding and cross-linking of GM-1 receptors. Although GM-1 binding is responsible for some of the effects of Etx and Ctx, it has been found that certain effects of the toxins, such as immunomodulation, are not mediated through GM-1 binding. The peptides of the invention are fragments of the beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as normal EtxB and CtxB subunits, except that they do not bind or cross link GM-1. They may be used in medicine as an immunomodulator or adjuvant. They may also be used as an inhibitor for toxin-induced diarrhoea. Therefore, the peptides may be used in the production of a composition for treating, preventing and/or modulating a disease associated with an immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463 represent preferred peptides of the invention, AAY87460 being particularly preferred.

CC Sequence 21 AA;

Query Match 100.0%; Score 107; DB 21; Length 21;
 Best Local Similarity 100.0%; Pred. No. 2.6e-11;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEFQVEVPSQHIDSQKAI 21
 ||||||||||||||||||
 DB 1 getfqvevpsqhidsqkai 21

RESULT 2

AA72545 AAR72545 standard; peptide; 93 AA.

XX AAR72545;

DT 28-NOV-1995 (first entry)

DE ADP-ribosylating toxin (verotoxin-1 B-subunit).

KM ADP-ribosylating toxin; pertussis holotoxin; B-subunit;

KW active site; E. coli heat labile toxin; verotoxin-1;

KM Bordetella pertussis vaccines.

XX Bacteria sp.

PN EP646599-A.

PD 05-APR-1995.

PF 23-AUG-1994; 94EP-0306219.

PR 24-AUG-1993; 93US-0110947.

PR 31-MAY-1994; 94US-0251121.

XX (CONN-) CONNAGHT LAB LTD.

PA (UYAL-) UNIV ALBERTA.

PI Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;

PI Oomen R, Read RJ, Stein PE;

XX WPI; 1995-132623/18.

XX New modified forms of pertussis holotoxin - developed using

XX crystalline forms of pertussis holotoxin and its complexes with

XX other molecules

PS Disclosure; Fig 5; 54pp; English.

XX AAR72540-R72545 are structurally equivalent B-subunits from three CC ADP-ribosylating toxins, pertussis holotoxin (PT), E. coli heat CC labile toxin (LT), and verotoxin-1 (VT). The structural CC information obtained from these comparisons was used to identify CC sites which contribute to PT's biological activity. By modifying CC these sites the claimed PT mutants of the invention were produced, CC they can be used in the development of vaccines against Bordetella CC pertussis infection.

CC Sequence 93 AA;

Query Match 100.0%; Score 107; DB 16; Length 93;
 Best Local Similarity 100.0%; Pred. No. 1.5e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEFQVEVPSQHIDSQKAI 21
 ||||||||||||||||||
 DB 35 getfqvevpsqhidsqkai 55

RESULT 3

AA71816 AAY1816 standard; peptide; 93 AA.

XX AAY1816;

DT 08-DEC-1999 (first entry)

DE *Escherichia coli* verotoxin-1 B-subunit.

KM ADP-ribosylating toxin; PT; crystalline pertussis holotoxin; ETA; DT;

KW three-dimensional structure; LT; immunoprotective; infection.

XX *Escherichia coli*.

PN US5965385-A.

PD 12-OCT-1999.

PF 06-JUN-1995; 95US-0467974.

PR 22-AUG-1994; 94US-0292968.

PR 24-AUG-1993; 93US-0110947.

PR 31-MAY-1994; 94US-0251121.

XX (CONN-) CONNAGHT LAB LTD.

PA (UYAL-) UNIV ALBERTA.

PI Read RJ, Stein PE, Cockle SA, Loosmore S, Klein MH, Armstrong GD;

PI Hazes B, Oomen RP;

XX WPI; 1999-579908/49.

XX New method for producing modified pertussis holotoxin -

XX Example 3; Fig 5; 41pp; English.

XX A method has been developed of producing a modified pertussis holotoxin, CC involving analysis of the 3-dimensional form of the crystalline CC holotoxin. The pertussis holotoxin modification process comprises: CC (1) identification of at least one amino acid (aa) residue of the CC holotoxin for modification by analysing the 3-dimensional form of the CC crystalline holotoxin, in relation to known information of the protein CC structure and function; (2) effecting mutagenesis (by removing or CC replacing a nucleotide sequence encoding at least one (aa)) of a tox CC operon; and (3) expressing mutant tox box in a *Bordetella* organism to CC produce the modified holotoxin. This method is used for modifying CC pertussis holotoxin, by studying its 3-dimensional crystalline CC structure. Modifying the holotoxin, alters its biological properties. CC By analysing the 3-dimensional crystalline structure of the pertussis

CC holotoxin, functional (aa) which affect biological properties of the
 CC pertussis holotoxin can be identified. This can be used to predict (aa)
 CC which contribute to the toxicity of the holotoxin to produce
 CC immunoprotective, genetically-detoxified analogues of pertussis
 CC holotoxin. The present sequence represents an ADP-ribosylating toxin
 CC B-subunit peptide used in the exemplification of the present
 CC invention.

XX Sequence 93 AA;

Query Match 100.0%; Score 107; DB 20; Length 93;
 Best Local Similarity 100.0%; Pred. No. 1.5e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSOHIDSOKKAI 21
 |||||
 Db 35 getfqvpgsqhidsqkai 55

RESULT 4

AAW95226 standard; peptide; 93 AA.

XX AAW95226;

DT 16-MAR-1999 (first entry)

DE E. coli heat-labile toxin (LT) beta-subunit sequence.

KW Pertussis holotoxin; PT: modified; effector; toxicity; cell binding;

KW enzyme; ADP-ribosylation; toxin; crystalline; X-ray crystallography;

KW structural analysis; interacting site; mitogenicity; adjuvantility;

heat-labile; LT.

XX Escherichia coli.

XX US856122-A.

PD 05-JAN-1999.

PF 22-AUG-1994; 94US-0292968.

PR 22-AUG-1994; 94US-0292968.

PR 24-AUG-1993; 93US-0110947.

PR 31-MAY-1994; 94US-0251121.

PA (UYAL-) UNIV ALBERTA.

PI Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;

PI Oomen RP, Read RJ, Stein PE;

DR WPI: 1999-105104/09.

XX Modifications to e.g. enzymatic activity, mitogenicity and cell

PT binding of pertussis holotoxin - by identifying interaction sites of

PT a molecule with crystalline toxin and modifying the identified site

XX Example 3; Fig 5; 40pp; English.

XX The invention relates to methods of preparing a pertussis holotoxin (PT)

CC having a modified biological activity. One method comprises identifying

CC at least 1 site in a PT that interacts with a molecule that is capable of

CC forming a complex with the holotoxin and which molecule is an effector

CC molecule which is an adenine nucleotide and which site contributes to

CC toxicity, cell binding or enzymatic activity of PT. The functional

CC toxins such as diphtheria toxin (DT), Pseudomonas exotoxin A (ETA), the
 CC heat-labile toxin of E. coli (LT) and verotoxin-1 (VT). The present
 CC sequence represents the beta-subunit of LT toxin.

XX Sequence 93 AA;

Query Match 100.0%; Score 107; DB 20; Length 93;
 Best Local Similarity 100.0%; Pred. No. 1.5e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSOHIDSOKKAI 21
 |||||
 Db 35 getfqvpgsqhidsqkai 55

RESULT 5

AAV68365 standard; peptide; 93 AA.

XX AAV68365;

DT 17-APR-2000 (first entry)

DE Heat labile toxin B subunit SEQ ID NO:26.

KW Pertussis toxin; heat labile toxin; verotoxin; exotoxin A; holotoxin;

KW diphtheria toxin; ADP-ribosylating toxin; mannose binding protein;

KW infection; crystal structure; X-ray crystallography; detoxification;

XX immunogenic.

XX Escherichia coli.

XX US6018022-A.

PD 25-JAN-2000.

PF 06-JUN-1995; 95US-0467976.

PR 22-AUG-1994; 94US-0292968.

PR 24-AUG-1993; 93US-0110947.

PR 31-MAY-1994; 94US-0251121.

PA (CONN-) CONNAUGHT LAB LTD.

PI (UYAL-) UNIV ALBERTA.

PI Read RJ, Stein PE, Cockle SA, Loosmore S, Klein MH, Armstrong GD;

PI Hazes B, Oomen RP;

DR WPI: 2000-136703/12.

XX Crystalline form of isolated pertussis holotoxin useful in studying

PT proteins which have functional resemblance -

XX Example 3; Fig 5; 42pp; English.

XX The present invention describes a crystalline form of isolated

CC pertussis holotoxin, in which the molecules of pertussis toxin have

CC a three dimensional structure represented in the specification,

CC complexed with a polysaccharide molecule capable of forming a complex

CC with the holotoxin. The crystalline form of the pertussis holotoxin

CC can be used in a comparison with other proteins which have functional

CC resemblance to pertussis holotoxin with the aim of modifying other

CC proteins. Identifying the unknown sites of toxicity by comparison

CC with the three dimensional structure of pertussis holotoxin provides a

CC technique for detoxification of toxins to produce useful immunogenic

CC but non-toxic analogues. It can also be used as a primary standard to

CC measure the quantity, purity or efficacy of less pure compositions

CC containing pertussis toxin. AAV68340 to AAV68385 represent peptides

CC used in the exemplification of the present invention.

XX Sequence 93 AA;

Query Match 100.0%; Score 107; DB 21; Length 93;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSCHIDSQKAI 21
DB 35 getfvevpgsqhidsqkai 55

RESULT 6
AAB6239
ID AAB6239 standard; Protein; 93 AA.

AC AAB6239;
DT 03-APR-2001 (first entry)
DE E coli verotoxin-1 B subunit SEQ ID NO: 26.

XX Pertussis toxin; crystal structure; whooping cough; biological activity;
KM lymphocytosis-promoting factor; histamine-sensitising factor;
KM islet-activating protein.

XX Escherichia coli.

XX US6168928-B1.

XX 02-JAN-2001.

XX 21-MAY-1998; 98US-0082514.

XX 22-AUG-1994; 94US-0292968.

XX 24-AUG-1993; 93US-0110947.

XX 31-MAY-1994; 94US-0251121.

XX (CONN-) CONNUGHT LAB LTD.

XX Read RJ, Cockle SA, Oomen RP, Loosmore S, Klein MH, Armstrong GD;
PI Hazen B, Stein PE;

XX WPI: 2001-122260/13.

XX Modifying pertussis holotoxin to produce detoxified PT analogs,
PT comprising analyzing crystalline structure of toxin, to identify sites
PT of toxicity, cell binding or enzyme activity of PT and modifying
PT identified site -

XX Example 3; Fig 5; 41pp; English.

XX The present invention provides a method for producing a pertussis toxin
CC (also designated lymphocytosis-promoting factor, histamine-sensitising
CC factor and islet activating protein) with a modified biological activity,
CC involving analysing the crystal structure of the protein to identify
CC active sites which can then be modified. This may lead to an alteration
CC in the toxicity, cell binding or enzyme activity of the toxin. This can
CC be used in the production of immunoprotective analogues of pertussis
CC toxin. Pertussis toxin is the cause of whooping cough following infection
CC by Bordetella pertussis.

XX Sequence 93 AA;

Query Match 100.0%; Score 107; DB 22; Length 93;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSCHIDSQKAI 21
DB 35 getfvevpgsqhidsqkai 55

RESULT 7

AAB62379
ID AAB62379 standard; Protein; 103 AA.

XX AAB62379;

DT 29-JUN-2001 (first entry)

XX E. coli LTB protein variant (GI: 494265).

XX ABB5B subunit protein; mutation; recombinant; cholera toxin B; CTB;
KM immunogen; autoimmune disease; neuron demyelinating disease; rCTB; LTB;
KM infection; parasitic; immunosuppressive; antitubercular; antineoplastic;
KM antidiabetic; neuroprotective; vaccine; heat labile toxin B; variant.

XX Escherichia coli.

XX WO200127144-A2.

XX 19-APR-2001.

XX 05-OCT-2000; 2000WO-US27607.

XX 08-OCT-1999; 99US-0158561.

XX (ACTI-) ACTIVE BIOTECH AB.

XX Handley HH, Haaparanta T, Ewalt KL;

XX WPI: 2001-281974/29.

XX Recombinant ABB5B subunit protein comprising a mutation that alters the
PT number residues available for chemical modification, useful for
PT covalently linking to an immunogen or vaccine which can be used for
PT treating autoimmune diseases -

XX Disclosure; Page 72-73; 78pp; English.

XX The invention relates to a recombinant ABB5B subunit protein comprising at
CC least one mutation, where the mutation alters the number of amino acid
CC residues available for chemical modification as compared to a wild type

CC ABB5B subunit protein, and where the recombinant protein retains an
CC effective target ligand binding affinity. A recombinant ABB5B subunit
CC protein such as cholera toxin B protein (CTB) can be specifically
CC covalently linked at lysines or cysteines to an immunogen or vaccine.
CC Recombinant mutant CTB (rCTB) can be used in the treatment of autoimmune
CC diseases e.g. rheumatoid arthritis, encephalomyelitis (or other neuron
CC demyelinating diseases) and diabetes. The rCTB or other B subunits of the
CC invention can also be used to induce tolerance to infection, e.g.
CC parasitic infection. The present sequence represents a variant of the
CC E. coli heat labile toxin B (LTB) protein.

XX Sequence 103 AA;

Query Match 100.0%; Score 107; DB 22; Length 103;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSCHIDSQKAI 21
DB 45 getfvevpgsqhidsqkai 65

RESULT 8
AAB62377
ID AAB62377 standard; Protein; 123 AA.

XX AAB62377;

DT 29-JUN-2001 (first entry)

XX E. coli LTB protein variant (GI: 223254).

KM AB5B subunit protein; mutation; recombinant; cholera toxin B; CTB;
 KM immunogen; autoimmune disease; neuron demyelinating disease; ICTB; LTB;
 KM infection; parasitic; immunosuppressive; antiarthritic; antineumatic;
 KM antidiabetic; neuroprotective; vaccine; heat labile toxin B; variant.
 OS Escherichia coli.
 XX WO200127144-A2.
 PN 19-APR-2001.
 XX
 PD 05-OCT-2000; 2000WO-US27607.
 XX
 PF 08-OCT-1999; 99US-0158561.
 XX
 PR (ACTI-) ACTIVE BIOTECH AB.
 PA
 XX Handley HH, Haaparanta T, Ewalt KL;
 PI WPI; 2001-281974/29.
 XX
 DR The invention relates to a recombinant AB5B subunit protein comprising at
 XX least one mutation, where the mutation alters the number of amino acid
 XX residues available for chemical modification as compared to a wild type
 XX AB5B subunit protein, and where the recombinant protein retains an
 XX effective target ligand binding affinity. A recombinant AB5B subunit
 XX protein such as cholera toxin B protein (CTB) can be specifically
 XX covalently linked at lysines or cysteines to an immunogen or vaccine.
 XX Recombinant mutant CTB (rCTB) can be used in the treatment of autoimmune
 XX diseases e.g. rheumatoid arthritis, encephalomyelitis (or other neuron
 XX demyelinating diseases) and diabetes. The rCTB or other B subunits of the
 XX invention can also be used to induce tolerance to infection, e.g.
 XX parasitic infection. The present sequence represents a variant of the
 XX E. coli heat labile toxin B (LTB) protein.
 SO Sequence 123 AA;

Query Match 100.0%; Score 107; DB 22; Length 123;
 Best Local Similarity 100.0%; Pred. No. 2.1e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPSQHIDSOKKAI 21
 Db 65 getfqvevpsqhdsgkai 85
 ||||||||||||||||
 ||||||||||||||||

RESULT 9
 AAB62373 standard; Protein; 124 AA.
 ID AAB62373;
 AC
 XX AAB62373;
 XX
 DT 29-JUN-2001 (first entry)
 XX
 DE E. coli LTB protein variant (GI: 1169505).
 XX
 KM AB5B subunit protein; mutation; recombinant; cholera toxin B; CTB;
 KM immunogen; autoimmune disease; neuron demyelinating disease; ICTB; LTB;
 KM infection; parasitic; immunosuppressive; antiarthritic; antineumatic;
 KM antidiabetic; neuroprotective; vaccine; heat labile toxin B; variant.
 XX Escherichia coli.
 OS
 XX
 PN WO200127144-A2.
 XX

PD 19-APR-2001.
 XX
 XX 05-OCT-2000; 2000WO-US27607.
 PF
 XX 08-OCT-1999; 99US-0158561.
 PR
 XX (ACTI-) ACTIVE BIOTECH AB.
 PA
 XX Handley HH, Haaparanta T, Ewalt KL;
 PI WPI; 2001-281974/29.
 XX
 DR Recombinant AB5B subunit protein comprising a mutation that alters the
 XX number residues available for chemical modification, useful for
 XX PT covalently linking to an immunogen or vaccine which can be used for
 XX PT treating autoimmune diseases -
 XX
 PS Disclosure; Page 70; 78pp; English.
 XX
 XX The invention relates to a recombinant AB5B subunit protein comprising at
 XX least one mutation, where the mutation alters the number of amino acid
 XX residues available for chemical modification as compared to a wild type
 XX AB5B subunit protein, and where the recombinant protein retains an
 XX effective target ligand binding affinity. A recombinant AB5B subunit
 XX protein such as cholera toxin B protein (CTB) can be specifically
 XX covalently linked at lysines or cysteines to an immunogen or vaccine.
 XX Recombinant mutant CTB (rCTB) can be used in the treatment of autoimmune
 XX diseases e.g. rheumatoid arthritis, encephalomyelitis (or other neuron
 XX demyelinating diseases) and diabetes. The rCTB or other B subunits of the
 XX invention can also be used to induce tolerance to infection, e.g.
 XX parasitic infection. The present sequence represents a variant of the
 XX E. coli heat labile toxin B (LTB) protein.
 SO Sequence 124 AA;

Query Match 100.0%; Score 107; DB 22; Length 124;
 Best Local Similarity 100.0%; Pred. No. 2.1e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPSQHIDSOKKAI 21
 Db 66 getfqvevpsqhdsgkai 86
 ||||||||||||||||
 ||||||||||||||||

RESULT 10
 AAB62375 standard; Protein; 124 AA.
 ID AAB62375;
 AC
 XX AAB62375;
 XX
 DT 29-JUN-2001 (first entry)
 XX
 DE E. coli LTB protein variant (GI: 145833).
 XX
 KM AB5B subunit protein; mutation; recombinant; cholera toxin B; CTB;
 KM immunogen; autoimmune disease; neuron demyelinating disease; ICTB; LTB;
 KM infection; parasitic; immunosuppressive; antiarthritic; antineumatic;
 KM antidiabetic; neuroprotective; vaccine; heat labile toxin B; variant.
 XX Escherichia coli.
 OS
 XX
 PN WO200127144-A2.
 XX
 PD 19-APR-2001.
 XX
 PF 05-OCT-2000; 2000WO-US27607.
 XX
 PR 08-OCT-1999; 99US-0158561.
 XX
 PA (ACTI-) ACTIVE BIOTECH AB.
 XX
 PI Handley HH, Haaparanta T, Ewalt KL;

XX
DR WPI: 2001-281974/29.
XX
XX Recombinant AB5B subunit protein comprising a mutation that alters the
PT number residues available for chemical modification, useful for
PT covalently linking to an immunogen or vaccine which can be used for
PT treating autoimmune diseases -
XX
PS Disclosure: Page 71; 78pp; English.
XX
XX The invention relates to a recombinant AB5B subunit protein comprising at
CC least one mutation, where the mutation alters the number of amino acid
CC residues available for chemical modification as compared to a wild type
CC AB5B subunit protein, and where the recombinant protein retains an
CC effective target ligand binding affinity. A recombinant AB5B subunit
CC covalently linked at lysines or cysteines to an immunogen or vaccine.
CC Recombinant mutant CTB (rCTB) can be used in the treatment of autoimmune
CC diseases e.g. rheumatoid arthritis, encephalomyelitis (or other neuron
CC demyelinating diseases) and diabetes. The rCTB or other B subunits of the
CC invention can also be used to induce tolerance to infection, e.g.
CC parasitic infection. The present sequence represents a variant of the
CC E. coli heat labile toxin B (LTB) protein.
XX
SQ Sequence 124 AA:

Query Match 100.0%; Score 107; DB 22; Length 124;
Best Local Similarity 100.0%; Pred. No. 2.1e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEFQVEVPGSQHDSQKAI 21
DB 66 getfgyevpgsqhdsqkai 86
|||||

RESULT 11
AAB62380
ID AAB62380 standard; Protein; 124 AA.
XX
AC AAB62380;
XX
DT 29-JUN-2001 (first entry)
XX
DE E. coli LTB protein variant (GI: 69630).
XX
XX AB5B subunit protein; mutation; recombinant; cholera toxin B; CTB;
KW immunogen; autoimmune disease; neuron demyelinating disease; rCTB; LTB;
KW infection; parasitic; immunosuppressive; antiarthritic; antirheumatic;
KW antidiabetic; neuroprotective; vaccine; heat labile toxin B; variant.
XX
OS Escherichia coli.
XX
PN WO200127144-A2.
XX
PD 19-APR-2001.
XX
PF 05-OCT-2000; 2000WO-US27607.
XX
PR 08-OCT-1999; 99US-0158561.
XX
PA (ACTI-) ACTIVE BIOTECH AB.
XX
PI Handley HH, Haaparanta T, Ewalt KL;
DR WPI: 2001-281974/29.
XX
XX Recombinant AB5B subunit protein comprising a mutation that alters the
PT number residues available for chemical modification, useful for
PT covalently linking to an immunogen or vaccine which can be used for
PT treating autoimmune diseases -
XX
PS Disclosure: Page 73; 78pp; English.

XX
CC The invention relates to a recombinant AB5B subunit protein comprising at
CC least one mutation, where the mutation alters the number of amino acid
CC residues available for chemical modification as compared to a wild type
CC AB5B subunit protein, and where the recombinant protein retains an
CC effective target ligand binding affinity. A recombinant AB5B subunit
CC protein such as cholera toxin B protein (CTB) can be specifically
CC covalently linked at lysines or cysteines to an immunogen or vaccine.
CC Recombinant mutant CTB (rCTB) can be used in the treatment of autoimmune
CC diseases e.g. rheumatoid arthritis, encephalomyelitis (or other neuron
CC demyelinating diseases) and diabetes. The rCTB or other B subunits of the
CC invention can also be used to induce tolerance to infection, e.g.
CC parasitic infection. The present sequence represents a variant of the
CC E. coli heat labile toxin B (LTB) protein.
XX
SQ Sequence 124 AA:

Query Match 100.0%; Score 107; DB 22; Length 124;
Best Local Similarity 100.0%; Pred. No. 2.1e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEFQVEVPGSQHDSQKAI 21
DB 66 getfgyevpgsqhdsqkai 86
|||||

RESULT 12
AAB73241
ID AAB73241 standard; Protein; 134 AA.
XX
AC AAB73241;
XX
DT 14-MAY-2001 (first entry)
XX
DE Recombinant exotoxin protein variant LTBP.
XX
XX Exotoxin mucosal cell binding motif; nucleic acid delivery;
KW nucleic acid affinity domain; heat-labile enterotoxin.
XX
OS Unidentified.
XX
PN WO200111960-A1.
XX
PD 22-FEB-2001.
XX
PF 18-AUG-2000; 2000WO-US22715.
XX
PR 18-AUG-1999; 99US-0149294.
XX
PA (AGRI-) AGRIVAX INC.
XX
PI Welter LM;
XX
DR WPI: 2001-211103/21.
XX
DR N-PSDB: AAF75712.
XX
XX Novel exotoxin protein variant useful as protein carrier for
PT facilitating gene delivery, comprises a mucosal cell binding motif of
PT an exotoxin and a nucleic acid affinity domain -
XX
PS Example 1; Fig 9; 57pp; English.
XX
XX The present invention relates to recombinant exotoxin protein variants,
CC which comprise an exotoxin mucosal cell binding motif and a nucleic acid
CC affinity domain. The present sequence is one such protein variant. In the
CC present invention the heat-labile enterotoxin (LT) of Escherichia coli
CC was used as an exotoxin. LT is a member of the ADP-ribosylating exotoxin
CC (babe) family. The protein variants are useful for selectively delivering
CC nucleic acid to mucosal cells, for inducing an immune response when the
CC nucleic acid encodes an antigen to which the immune response is desired,
CC for selectively delivering a gene to a mucosal cell, and for achieving
CC expression of a protein in a subject, by administering a composition

CC comprising the protein variant.
XX
SQ Sequence 134 AA;

Query Match 100.0%; Score 107; DB 22; Length 134;
Best Local Similarity 100.0%; Pred. No. 2.3e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPSQHDSQKAI 21
|
66 getfvevpsqhdsgkai 86

RESULT 13

AAB73242 ID AAB73242 standard; Protein: 142 AA.

XX AAB73242;

XX 14-MAY-2001 (first entry)

DE Recombinant exotoxin protein variant LTBpLh.

XX Exotoxin mucosal cell binding motif; nucleic acid delivery;
KM nucleic acid affinity domain; heat-labile enterotoxin.

XX Unidentified.

XX WO200111960-A1.

XX 22-FEB-2001.

XX 18-AUG-2000; 2000WO-US22715.

XX 18-AUG-1999; 99US-0149294.

XX (AGRI-) AGRIVAX INC.

XX Welter LM;

XX WPI: 2001-211103/21.

XX N-PSDB; AAF75713.

PT Novel exotoxin protein variant useful as protein carrier for
PT facilitating gene delivery, comprises a mucosal cell binding motif of
an exotoxin and a nucleic acid affinity domain

XX Example 1; Fig 10; 57pp; English.

XX The present invention relates to recombinant exotoxin protein variants,
CC which comprise an exotoxin mucosal cell binding motif and a nucleic acid
CC affinity domain. The present sequence is one such protein variant. In the
CC present invention the heat-labile enterotoxin (LT) of Escherichia coli
CC was used as an exotoxin. LT is a member of the ADP-ribosylating exotoxin
CC (bAR) family. The protein variants are useful for selectively delivering
CC nucleic acid to mucosal cells, for inducing an immune response when the
CC nucleic acid encodes an antigen to which the immune response is desired,
CC for selectively delivering a gene to a mucosal cell, and for achieving
CC expression of a protein in a subject, by administering a composition
CC comprising the protein variant.

XX Sequence 142 AA;

Query Match 100.0%; Score 107; DB 22; Length 142;

Best Local Similarity 100.0%; Pred. No. 2.5e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPSQHDSQKAI 21
|
66 getfvevpsqhdsgkai 86

RESULT 14

AAB73243 ID AAB73243 standard; Protein: 155 AA.

XX AAB73243;

XX 14-MAY-2001 (first entry)

DE Recombinant exotoxin protein variant LTB-P.

XX Exotoxin mucosal cell binding motif; nucleic acid delivery;
KM nucleic acid affinity domain; heat-labile enterotoxin.

XX Unidentified.

XX WO200111960-A1.

XX 22-FEB-2001.

XX 18-AUG-2000; 2000WO-US22715.

XX 18-AUG-1999; 99US-0149294.

XX (AGRI-) AGRIVAX INC.

XX Welter LM;

XX WPI: 2001-211103/21.

XX N-PSDB; AAF75714.

PT Novel exotoxin protein variant useful as protein carrier for
PT facilitating gene delivery, comprises a mucosal cell binding motif of
an exotoxin and a nucleic acid affinity domain

XX Example 1; Fig 11; 57pp; English.

XX The present invention relates to recombinant exotoxin protein variants,
CC which comprise an exotoxin mucosal cell binding motif and a nucleic acid
CC affinity domain. The present sequence is one such protein variant. In the
CC present invention the heat-labile enterotoxin (LT) of Escherichia coli
CC was used as an exotoxin. LT is a member of the ADP-ribosylating exotoxin
CC (bAR) family. The protein variants are useful for selectively delivering
CC nucleic acid to mucosal cells, for inducing an immune response when the
CC nucleic acid encodes an antigen to which the immune response is desired,
CC for selectively delivering a gene to a mucosal cell, and for achieving
CC expression of a protein in a subject, by administering a composition
CC comprising the protein variant.

XX Sequence 155 AA;

Query Match 100.0%; Score 107; DB 22; Length 155;

Best Local Similarity 100.0%; Pred. No. 2.8e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPSQHDSQKAI 21
|
66 getfvevpsqhdsgkai 86

RESULT 15

AAB73244 ID AAB73244 standard; Protein: 163 AA.

XX AAB73244;

XX 14-MAY-2001 (first entry)

DE Recombinant exotoxin protein variant LTB-P.

XX Exotoxin mucosal cell binding motif; nucleic acid delivery;
KM nucleic acid affinity domain; heat-labile enterotoxin; ds.

XX Unidentified.
 OS
 XX
 PN WO20011960-A1.
 XX
 PD 22-FEB-2001.
 XX
 PF 18-AUG-2000; 2000WO-US22715.
 XX
 PR 18-AUG-1999; 99US-0149294.
 XX
 PA (AGRI-) AGRIVAX INC.
 XX
 PI Welter LM;
 XX
 DR WPI; 2001-211103/21.
 XX N-PSDB; AAF5715.
 PT Novel exotoxin protein variant useful as protein carrier for
 PT facilitating gene delivery, comprises a mucosal cell binding motif of
 PT an exotoxin and a nucleic acid affinity domain
 XX
 PS Example 1; Fig 12; 57pp; English.
 CC
 CC The present invention relates to recombinant exotoxin protein variants,
 CC which comprise an exotoxin mucosal cell binding motif and a nucleic acid
 CC affinity domain. The present sequence is one such protein variant. In the
 CC present invention the heat-labile enterotoxin (LT) of Escherichia coli
 CC was used as an exotoxin. LT is a member of the ADP-ribosylating exotoxin
 CC (bare) family. The protein variants are useful for selectively delivering
 CC nucleic acid to mucosal cells, for inducing an immune response when the
 CC nucleic acid encodes an antigen to which the immune response is desired,
 CC for selectively delivering a gene to a mucosal cell, and for achieving
 CC expression of a protein in a subject, by administering a composition
 CC comprising the protein variant.
 SQ Sequence 163 AA;

Query Match 100.0%; Score 107; DB 22; Length 163;
 Best Local Similarity 100.0%; Pred. No. 3e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GETFOVEVPGSOHIDSQKRAI 21
 ||||||||||||||||||
 Db 66 getfgvevpsqhdsgkrai 86

Search completed: July 3, 2002, 08:54:57
 Job time: 287 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 08:55:39 ; Search time 34.88 Seconds
(without alignments)
14.706 Million cell updates/sec

Title: US-09-786-648-5

Perfect score: 107

Sequence: 1 GETFQVEVPGSHIDSOKKAI 21

Scoring table: BLOSUM62

Searched: Gapop 10.0, Gapext 0.5

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5E.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	100.0	93	2	US-08-292-968-26
2	107	100.0	93	2	US-08-467-974-26
3	107	100.0	93	2	US-08-467-536-26
4	107	100.0	93	3	US-08-467-976-26
5	107	100.0	93	4	US-09-082-514-26
6	107	100.0	371	2	US-08-829-026A-6
7	101	94.4	102	3	US-08-952-337-5
8	101	94.4	102	3	US-08-952-337-6
9	101	94.4	103	2	US-08-472-171-2
10	101	94.4	103	2	US-08-894-526-2
11	101	94.4	103	2	US-09-013-047-2
12	101	94.4	103	4	US-09-374-597-2
13	101	94.4	103	4	US-09-191-852-21
14	101	94.4	103	5	PCT-US95-13376-21
15	101	94.4	133	3	US-08-952-337-1
16	101	94.4	123	3	US-08-952-337-2
17	101	94.4	124	2	US-08-747-410-2
18	89	83.2	124	1	US-08-449-045C-4
19	89	83.2	124	2	US-08-435-605A-12
20	89	83.2	124	6	5223610-3
21	54	50.5	448	4	US-08-878-989-2
22	54	50.5	448	4	US-09-272-796-2
23	54	50.5	508	4	US-09-344-700-4
24	43	40.2	1810	5	PCT-US95-11684-4
25	42	39.3	855	5	US-09-027-337-2
26	41	38.3	1022	1	US-08-271-364A-8
27	41	38.3	1022	2	US-08-222-715B-27

28	40.5	37.9	856	3	US-08-709-784-2	Sequence 2, Appl
29	40.5	37.9	862	2	US-08-209-521-23	Sequence 23, Appl
30	40.5	37.9	862	2	US-08-209-521-30	Sequence 30, Appl
31	40.5	37.9	862	4	US-09-059-461-2	Sequence 2, Appl
32	40.5	37.9	862	4	US-08-961-810-133	Sequence 133, App
33	40.5	37.9	862	4	US-08-352-902D-133	Sequence 1, Appl
34	40	37.4	774	3	US-08-902-632-2	Sequence 2, Appl
35	40	37.4	774	3	US-09-073-354-1	Sequence 1, Appl
36	40	37.4	774	3	US-08-656-005A-1	Sequence 1, Appl
37	40	37.4	774	4	US-09-073-259-1	Sequence 1, Appl
38	40	37.4	774	4	US-09-363-095-1	Sequence 1, Appl
39	40	37.4	774	4	US-09-418-027-1	Sequence 1, Appl
40	40	37.4	778	2	US-08-906-925-4	Sequence 4, Appl
41	40	37.4	779	5	US-08-375-134-12	Sequence 12, Appl
42	40	37.4	779	5	PCT-US95-15263-12	Sequence 12, Appl
43	39.5	36.9	1090	4	US-09-346-237-5	Sequence 5, Appl
44	39	36.4	75	1	US-08-350-884-35	Sequence 35, Appl
45	39	36.4	75	1	US-08-709-173-35	Sequence 35, Appl

ALIGNMENTS

RESULT 1
US-08-292-968-26
Sequence 26, Application US/08292968
Patent No. 5856122
GENERAL INFORMATION:
APPLICANT: READ, Randy J.
APPLICANT: STEIN, Penelope E.
APPLICANT: COCKLE, Stephen A.
APPLICANT: OOMEN, Raymond P.
APPLICANT: LOOSMORE, Sheena
APPLICANT: KLEIN, Michel H.
APPLICANT: ARMSTRONG, Glen D.
APPLICANT: HAZES, Bart
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Slim & Mcburney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/292,968
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/110,947
FILING DATE: 24-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/251,121
FILING DATE: 31-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-388
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-292-968-26

Query Match 100.0%; Score 107; DB 2; Length 93;
Best Local Similarity 100.0%; Pred. No. 2.7e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSOHIDSOKKAI 21
|||||

DB 35 GETFOVEVPGSOHIDSOKKAI 55

RESULT 2

US-08-467-974-26
; Sequence 26, Application US/08467974
; Patent No. 5965385

GENERAL INFORMATION:

APPLICANT: READ, Randy J.

APPLICANT: STEIN, Penelope E.

APPLICANT: COCKLE, Stephen A.

APPLICANT: OOMEN, Raymond P.

APPLICANT: LOOSMORE, Sheena

APPLICANT: KLEIN, Michael H.

APPLICANT: ARMSTRONG, Glen D.

APPLICANT: HAZES, Bart

TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN

NUMBER OF SEQUENCES: 46

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney

STREET: Suite 701, 330 University Avenue

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/467,974

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/467,536

FILING DATE: 06-JUN-1995

APPLICATION NUMBER: US 08/292,968

FILING DATE: 22-AUG-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/251,121

FILING DATE: 31-MAY-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/110,947

FILING DATE: 24-AUG-1993

ATTORNEY/AGENT INFORMATION:

NAME: STEWART, Michael I.

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-454 MIS:vg

TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 93 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-467-974-26

QY 1 GETFOVEVPGSOHIDSOKKAI 21
|||||

DB 35 GETFOVEVPGSOHIDSOKKAI 55

RESULT 3

US-08-467-536-26
; Sequence 26, Application US/08467536
; Patent No. 5977304

GENERAL INFORMATION:

APPLICANT: READ, Randy J.

APPLICANT: STEIN, Penelope E.

APPLICANT: COCKLE, Stephen A.

APPLICANT: OOMEN, Raymond P.

APPLICANT: LOOSMORE, Sheena

APPLICANT: KLEIN, Michael H.

APPLICANT: ARMSTRONG, Glen D.

APPLICANT: HAZES, Bart

TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN

NUMBER OF SEQUENCES: 46

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney

STREET: Suite 701, 330 University Avenue

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/467,536

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/292,968

FILING DATE: 22-AUG-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/251,121

FILING DATE: 31-MAY-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/110,947

FILING DATE: 24-AUG-1993

ATTORNEY/AGENT INFORMATION:

NAME: STEWART, Michael I.

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-455 MIS:vg

TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 93 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-467-536-26

Query Match 100.0%; Score 107; DB 2; Length 93;
Best Local Similarity 100.0%; Pred. No. 2.7e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSOHIDSOKKAI 21
|||||

DB 35 GETFOVEVPGSOHIDSOKKAI 55

RESULT 4

US-08-467-976-26

; Sequence 26, Application US/08467976
; Patent No. 6018022
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J.
; APPLICANT: STEIN, Penelope E.
; APPLICANT: COCKLE, Stephen A.
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: KLEIN, Michel H.
; APPLICANT: ARMSTRONG, Glen D.
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; City: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,976
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,968
; FILING DATE: 22-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/251,121
; FILING DATE: 31-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/110,947
; FILING DATE: 24-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-453 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-467-976-26

Query Match 100.0%; Score 107; DB 3; Length 93;
Best Local Similarity 100.0%; Pred. No. 2.7e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSQHDSOKKAI 21
|||||
Db 35 GETFOVEVPGSQHDSOKKAI 55

RESULT 5
US-09-082-514-26
; Sequence 26, Application US/09082514
; Patent No. 6168928
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J.
; APPLICANT: STEIN, Penelope E.
; APPLICANT: COCKLE, Stephen A.
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.

; APPLICANT: ARMSTRONG, Glen D.
; APPLICANT: HAZES, Bart
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; City: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/082,514
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,968
; FILING DATE: 24-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-810
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-082-514-26

Query Match 100.0%; Score 107; DB 4; Length 93;
Best Local Similarity 100.0%; Pred. No. 2.7e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSQHDSOKKAI 21
|||||
Db 35 GETFOVEVPGSQHDSOKKAI 55

RESULT 6
US-08-829-026A-6
; Sequence 6, Application US/08829026A
; Patent No. 5837825
; GENERAL INFORMATION:
; APPLICANT: Melnersmann, Richard J.
; APPLICANT: Khoury, Christian A.
; TITLE OF INVENTION: Campylobacter Jejuni Flagellin-Escherichia Coli LT-B Fusion
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Janelle S. Graeter
; STREET: Room 411, Building 005, BARC-W
; City: Beltsville
; STATE: MD
; COUNTRY: USA
; ZIP: 20705
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/829,026A
; FILING DATE: 18-AUG-1997
; CLASSIFICATION: 435

```

? ATTORNEY/AGENT INFORMATION:
?   NAME: Graeter, Janelle S.
?   REGISTRATION NUMBER: 35,024
?   REFERENCE/DOCKET NUMBER: 0106.97
? TELECOMMUNICATION INFORMATION:
?   TELEPHONE: 301-504-5676
?   TELEFAX: 301-504-5060
? INFORMATION FOR SEQ ID NO: 6:
?   SEQUENCE CHARACTERISTICS:
?     LENGTH: 371 amino acids
?     TYPE: amino acid
?     TOPOLOGY: linear
? MOLECULE TYPE: protein
?
US-08-629-026A-6

```

Query Match	100.0%;	Score 107;	DB 2;	Length 371;
Best Local Similarity	100.0%;	Pred. No. 1.5e-10;		
Matches 21; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

```
QY 1 GETFOVEVPGSQHIDSQKAI 21
    |||||
Db 49 GETFOVEVPGSQHIDSQKAI 69
```

```

RESULT 7
US-08-952-337-5
Sequence 5 / Application US/08952337
Patent No. 6019973
GENERAL INFORMATION:
APPLICANT: Holmgren, Jan
APPLICANT: Lebens, Michael R.
TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
FILE REFERENCE: 3846/00758
CURRENT APPLICATION NUMBER: US/08/952,337
CURRENT FILING DATE: 1998-01-05
EARLIER APPLICATION NUMBER: PCT/SE96/00570
EARLIER FILING DATE: 1996-05-02
EARLIER APPLICATION NUMBER: SE 9501682-0
EARLIER FILING DATE: 1995-05-05
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5
LENGTH: 102
TYPE: PRT
ORGANISM: Vibrio cholerae
US-08-952-337-5

```

Query Match	94.48;	Score 101;	DB 3;	Length 102;
Best Local Similarity	95.28;	Pred. No. 3.3e-10;		
Matches 20; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0

```

Oy      1 GETFQVEVPGSQHIDSQKKA 21
          | | | | | | | | | | | | |
Db      44 GATFQVEVPGSQHIDSQKKA 64

```

RESULT 8
 US-08-952-337-6
 : Sequence 6, Application US/08952337
 : Patient No. 6019973
 : GENERAL INFORMATION:
 : APPLICANT: Holmgren, Jan
 : APPLICANT: Lebens, Michael R.
 : TITLE OF INVENTION: HYDRO MOLECULES BETWEEN HEAT-LABILE
 : TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS
 : FILE REFERENCE: 3846/00758
 : CURRENT APPLICATION NUMBER: US/08/952, 337
 : CURRENT FILING DATE: 1998-01-05
 : EARLIER APPLICATION NUMBER: PCT/SE96/00570
 : EARLIER FILING DATE: 1996-05-02

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

query match	94.48;	Score 101;	DB 3;	Length 102;
Best Local Similarity	95.28;	Pred. No. 3.3e-10;		
Matches 20;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

```
QY 1 GETFOVEVPGSQHIDSQKAI 21
    | | | | | | | | | | | | | |
Db 44 GATFOVEVPGSQHIDSQKAI 64
```

```

RESULT          9
US-08-472-171-2
? Sequence 2, Application US/08472171
? Patent No. 5932714
?
? GENERAL INFORMATION:
? APPLICANT:  Loosmore, Sheena M.
? APPLICANT:  Yacoub, Reza K.
? APPLICANT:  Zealey, Gavin R.
? APPLICANT:  Klein, Michel H.
?
? TITLE OF INVENTION:  Expression Of Gene Products From
?                       Genetically Manipulated Strains Of Bordetella
?
? NUMBER OF SEQUENCES:  56
?
? CORRESPONDENCE ADDRESS:
? ADDRESSEE:  Slim & McBurney
? STREET:     330 University Avenue, Suite 701
? CITY:       Toronto
? STATE:      Ontario
? COUNTRY:    Canada
? ZIP:        M5G 1R7
?
? COMPUTER READABLE FORM:
?
? MEDIUM TYPE:  Floppy disk
?
? COMPUTER:  IBM PC compatible
? OPERATING SYSTEM:  PC-DOS/MS-DOS
? SOFTWARE:  PatentIn Release #1.0, Version #1.30
?
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER:  US/08/472,171
? FILING DATE:  07-JUN-1995
?
? CLASSIFICATION:  435
?
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:  US 08/393,334
? FILING DATE:  23-FEB-1995
?
? ATTORNEY/AGENT INFORMATION:
?
? NAME:  Stewart, Michael I.
? REGISTRATION NUMBER:  24,973
? REFERENCE/DOCKET NUMBER:  1038-507 MIS:vg
? TELECOMMUNICATION INFORMATION:
? TELEPHONE:  416-595-1155
? TELEFAX:   416-595-1163
? TELEX:     065-24567 Slnbas
?
? INFORMATION FOR SEQ ID NO:  2:
?
? SEQUENCE CHARACTERISTICS:
? LENGTH:  103 amino acids
? TYPE:    amino acid
? TOPOLOGY:  linear
?
? MOLECULE TYPE:  protein
?
? US-08-472-171-2

```

Query Match	94.48;	Score 101;	DB 2;	Length 103;
Best Local Similarity	95.28;	Pred. No. 3.3e-10;		
Matches 20;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

QY 1 GETFOVEVPGSQHIDSQKKAI 21

Db 45 GATFOVEYPSQSHIDSOKKAI 65

RESULT 10

US-08-894-526-2

; Sequence 2, Application US/08894526

; Patent No. 5942418

GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena M

APPLICANT: Yacoub, Reza K

APPLICANT: Zealey, Gavin R

APPLICANT: Klein, Michel H

TITLE OF INVENTION: EXPRESSION OF GENE PRODUCTS FROM

TITLE OF INVENTION: GENETICALLY MANIPULATED STRAINS OF BORDETELLA

NUMBER OF SEQUENCES: 57

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney

STREET: 6th Floor, 330 University Avenue

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/894,526

FILING DATE: 01-DEC-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-724 MIS:jb

TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 103 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-894-526-2

Query Match 94.4%; Score 101; DB 2; Length 103;

Best Local Similarity 95.2%; Pred. No. 3.3e-10;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETFOVEYPSQSHIDSOKKAI 21

Db 45 GATFOVEYPSQSHIDSOKKAI 65

RESULT 11

US-09-013-047-2

; Sequence 2, Application US/09013047

; Patent No. 5998168

GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena M.

APPLICANT: Yacoub, Reza K.

APPLICANT: Zealey, Gavin R.

APPLICANT: Klein, Michel H.

TITLE OF INVENTION: Expression of Gene Products From

TITLE OF INVENTION: Genetically Manipulated Strains of Bordetella

NUMBER OF SEQUENCES: 56

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney

STREET: 330 University Avenue, 6th Floor

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/013,047

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/472,171

FILING DATE: 07-JUN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/393,334

FILING DATE: 23-FEB-1995

ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I.

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-507 MIS:vg

TELECOMMUNICATION INFORMATION:

TELEPHONE: 416-595-1155

TELEFAX: 416-595-1163

TELEX: 065-24567 Simbas

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 103 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-013-047-2

Query Match 94.4%; Score 101; DB 2; Length 103;

Best Local Similarity 95.2%; Pred. No. 3.3e-10;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETFOVEYPSQSHIDSOKKAI 21

Db 45 GATFOVEYPSQSHIDSOKKAI 65

RESULT 12

US-09-374-597-2

; Sequence 2, Application US/09374597

; Patent No. 6140082

GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena M.

APPLICANT: Yacoub, Reza K.

APPLICANT: Zealey, Gavin R.

APPLICANT: Klein, Michel H.

TITLE OF INVENTION: Expression of Gene Products From

TITLE OF INVENTION: Genetically Manipulated Strains of Bordetella

NUMBER OF SEQUENCES: 56

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney

STREET: 330 University Avenue, 6th Floor

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/374,597

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/393,334
FILING DATE: FEBRUARY 23, 1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24973
REFERENCE/DOCKET NUMBER: 1038-964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 416-595-1155
TELEFAX: 416-595-1163
TELEX: 065-24567 Simbas
INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-374-597-2

Query Match 94.4%; Score 101; DB 4; Length 103;
Best Local Similarity 95.2%; Pred. No. 3.3e-10;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSOHIDSOKKAI 21
DB 45 GATFOVEVPGSOHIDSOKKAI 65

RESULT 13
US-09-191-852-21
Sequence 21, Application US/09191852
GENERAL INFORMATION:
APPLICANT: Charles J. Arntzen, Hugh S. Mason, and Tariq A. Haq
TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski L.L.P.
STREET: 1301 McKinney, Suite 5100
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77010
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/191,852
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13376
FILING DATE: 24-OCT-1995
APPLICATION NUMBER: 08/817,906
FILING DATE: 04-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Fox, David L.
REGISTRATION NUMBER: 40,612
REFERENCE/DOCKET NUMBER: P015900S1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-651-5151
TELEFAX: 713-651-5246
INFORMATION FOR SEQ. ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-09-191-852-21

Query Match 94.4%; Score 101; DB 4; Length 103;
Best Local Similarity 95.2%; Pred. No. 3.3e-10;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSOHIDSOKKAI 21
DB 45 GATFOVEVPGSOHIDSOKKAI 65

RESULT 14
PCT-US95-13376-21
Sequence 21, Application PC/TUS9513376
GENERAL INFORMATION:
APPLICANT: The Texas A&M University System
APPLICANT: 310 Wisenbaker
APPLICANT: College Station, Texas 77843-3369
TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pravel, Hewitt, Kimball & Krieger
STREET: 1177 West Loop South, 10th Floor
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77027-9095
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13376
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/328,716
FILING DATE: 24-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Jones, John W.
REGISTRATION NUMBER: 31,380
REFERENCE/DOCKET NUMBER: 36170/3P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-850-0165
TELEFAX: 713-850-0909
INFORMATION FOR SEQ. ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear
PCT-US95-13376-21

Query Match 94.4%; Score 101; DB 5; Length 103;
Best Local Similarity 95.2%; Pred. No. 3.3e-10;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSOHIDSOKKAI 21
DB 45 GATFOVEVPGSOHIDSOKKAI 65

RESULT 15
US-08-952-337-1
Sequence 1, Application US/08952337
Patent No. 6019973
GENERAL INFORMATION:
APPLICANT: Holmgren, Jan
APPLICANT: Lebens, Michael R.
TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
FIDE REFERENCE: 3846/OD758
CURRENT APPLICATION NUMBER: US/08/952,337
CURRENT FILING DATE: 1998-01-05

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 09:03:21 ; Search time 401.04 Seconds
(without alignments)
18.431 Million cell updates/sec

Title: US-09-786-648-5

Perfect score: 107

Sequence: 1 GERTFYVEFGSHIDSOKKAI 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents AA Main:*

- 1: /cgn2_6/ptodata/2/paa/PCNUS_COMB.pep.*
- 2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
- 3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
- 4: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
- 5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
- 6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/2/paa/US097_COMB.pep.*
- 22: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
- 23: /cgn2_6/ptodata/2/paa/US099_COMB.pep.*
- 24: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
- 25: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
- 26: /cgn2_6/ptodata/2/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	100.0	21	21	US-09-786-648-5
2	107	100.0	93	5	US-08-110-947-10
3	107	100.0	93	5	US-08-110-947A-26
4	107	100.0	93	6	US-08-251-121-26
5	107	100.0	371	5	US-08-150-305A-3
6	107	100.0	371	11	US-08-784-218-6
7	107	100.0	371	12	US-08-829-026-5

8	101	94.4	21	21	US-09-786-648-4	Sequence 4, Appli
9	101	94.4	103	7	US-08-393-334-2	Sequence 2, Appli
10	101	94.4	103	11	US-08-782-832-15	Sequence 15, Appli
11	101	94.4	103	12	US-08-817-906-21	Sequence 21, Appli
12	101	94.4	103	22	US-09-836-433-14	Sequence 14, Appli
13	101	94.4	116	22	US-09-836-433-20	Sequence 20, Appli
14	101	94.4	119	22	US-09-836-433-22	Sequence 22, Appli
15	101	94.4	124	18	PCT-US99-30747-55	Sequence 55, Appli
16	101	94.4	124	18	US-09-470-124-55	Sequence 55, Appli
17	101	94.4	138	13	US-08-914-479A-2	Sequence 2, Appli
18	101	94.4	138	13	US-08-914-479A-2	Sequence 2, Appli
19	101	94.4	313	21	US-09-756-983-15	Sequence 15, Appli
20	101	94.4	351	21	US-09-756-983-18	Sequence 18, Appli
21	101	94.4	364	21	US-09-756-983-22	Sequence 22, Appli
22	95	88.8	124	1	PCT-US99-30747-57	Sequence 57, Appli
23	95	88.8	124	21	US-09-470-124-57	Sequence 57, Appli
24	89	83.2	124	21	US-09-760-234-7	Sequence 7, Appli
25	89	83.2	382	1	PCT-US01-08582-3	Sequence 3, Appli
26	89	83.2	382	1	PCT-US01-08582-4	Sequence 4, Appli
27	89	83.2	461	14	US-09-051-315-2	Sequence 2, Appli
28	89	83.2	461	18	US-09-423-493-2	Sequence 2, Appli
29	89	83.2	750	18	US-09-402-100-2	Sequence 2, Appli
30	89	83.2	1338	18	US-09-402-100-4	Sequence 4, Appli
31	76	71.0	15	11	US-08-732-371-1	Sequence 1, Appli
32	76	71.0	15	11	US-08-732-371A-1	Sequence 1, Appli
33	62	57.9	12	21	US-09-786-648-3	Sequence 3, Appli
34	54	50.5	286	21	US-09-758-445-415	Sequence 415, App
35	54	50.5	448	21	US-09-769-970-2	Sequence 2, Appli
36	54	50.5	508	19	US-09-563-997-4	Sequence 4, Appli
37	49	45.8	248	18	US-09-451-330-388	Sequence 388, App
38	49	45.8	267	17	US-09-391-631-3104	Sequence 3104, Ap
39	49	45.8	267	20	US-09-689-980-802	Sequence 802, App
40	49	45.8	270	17	US-09-391-631-3103	Sequence 3103, Ap
41	49	45.8	270	20	US-09-689-980-801	Sequence 801, App
42	48	44.9	340	16	US-09-252-691-6923	Sequence 6923, Ap
43	48	44.9	340	16	US-09-252-691C-6923	Sequence 6923, Ap
44	48	44.9	484	16	US-09-252-991A-29252	Sequence 29252, A
45	47	43.9	369	21	US-09-708-427-58248	Sequence 58248, A

ALIGNMENTS

US-09-786-648-5
; Sequence 5, Application US/09786648
; GENERAL INFORMATION:
; APPLICANT: Williams, Neil Andrew
; TITLE OF INVENTION: First, Timothy Raymond
; TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vacci
; FILE REFERENCE: 7438
; CURRENT APPLICATION NUMBER: US/09/786,648
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: PCT/GB99/02970
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: MS DOS
; SEQ ID NO 5
; LENGTH: 21
; TYPE: PRT
; ORGANISM: E. coli
; FEATURE:
; LOCATION: 45...65
; OTHER INFORMATION: Isolated or synthetic Etbx beta4-alpha2 loop fragment derivabl
; OTHER INFORMATION: porcine E. coli
; US-09-786-648-5

Query Match 100.0%, Score 107; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.4e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

NAME: STEWART, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-335
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-251-121-26

Query Match 100.0%; Score 107; DB 6; Length 93;
Best Local Similarity 100.0%; Pred. No. 9.3e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSOHIDSOKKAI 21
|||||
Db 35 GETFOVEVPGSOHIDSOKKAI 55

RESULT 5
US-08-150-305A-3
Sequence 3, Application US/08150305A
GENERAL INFORMATION:
APPLICANT: Meinersmann, Richard J.
TITLE OF INVENTION: CAMPYLOBACTER JEJUNI
TITLE OF INVENTION: FLAGELLIN-ESCHERICHIA COLI ST-B FUSION PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Janelle S. Graeter
STREET: Room 413, Building 005, BARC-W
CITY: Beltsville
STATE: MD
COUNTRY: USA
ZIP: 20705
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/150,305A
FILING DATE: 12-NOV-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Graeter, Janelle S.
REGISTRATION NUMBER: 35,024
REFERENCE/DOCKET NUMBER: 0431.92
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-504-5676
TELEFAX: 202-504-5060
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 371 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-150-305A-3

Query Match 100.0%; Score 107; DB 5; Length 371;
Best Local Similarity 100.0%; Pred. No. 5.5e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSOHIDSOKKAI 21
|||||
Db 49 GETFOVEVPGSOHIDSOKKAI 69

RESULT 6
US-08-784-218-6
Sequence 6, Application US/08784218
GENERAL INFORMATION:
APPLICANT: Meinersmann, Richard J.
TITLE OF INVENTION: CAMPYLOBACTER JEJUNI FLAGELLIN-ESCHERICHIA COLI LT-B FUSION
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Janelle S. Graeter
STREET: Room 411, Building 005, BARC-W
CITY: Beltsville
STATE: MD
COUNTRY: USA
ZIP: 20705
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/784,218
FILING DATE: 16-JAN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Graeter, Janelle S.
REGISTRATION NUMBER: 35,024
REFERENCE/DOCKET NUMBER: 0043.97
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-504-5676
TELEFAX: 301-504-5060
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 371 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-784-218-6

Query Match 100.0%; Score 107; DB 11; Length 371;
Best Local Similarity 100.0%; Pred. No. 5.5e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSOHIDSOKKAI 21
|||||
Db 49 GETFOVEVPGSOHIDSOKKAI 69

RESULT 7
US-08-829-026-5
Sequence 5, Application US/08829026
GENERAL INFORMATION:
APPLICANT: Meinersmann, Richard J.
TITLE OF INVENTION: CAMPYLOBACTER JEJUNI
TITLE OF INVENTION: FLAGELLIN-ESCHERICHIA COLI LT-B FUSION PROTEIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Janelle S. Graeter
STREET: Room 411, Building 005, BARC-W
CITY: Beltsville
STATE: MD
COUNTRY: USA
ZIP: 20705
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/829,026

FILING DATE: 18-AUG-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Graeter, Janelle S.
REGISTRATION NUMBER: 35,024
REFERENCE/DOCKET NUMBER: 0106.97
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-504-5676
TELEFAX: 301-504-5060
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 371 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-829-026-5

Query Match
Best Local Similarity 100.0%; Score 107; DB 12; Length 371;
Pred. No. 5.5e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSOHIDSQKAI 21
|||||
DB 49 GETFOVEVPGSOHIDSQKAI 69

RESULT 8
US-09-786-648-4
Sequence 4, Application US/09786648
GENERAL INFORMATION:
APPLICANT: Williams, Neil Andrew
TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccine
FILE REFERENCE: 7438
CURRENT APPLICATION NUMBER: US/09/786,648
PRIOR FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: PCT/GB99/02970
PRIOR FILING DATE: 1999-09-07
NUMBER OF SEQ ID NOS: 6
SOFTWARE: MS DOS
SEQ ID NO: 4
LENGTH: 21
TYPE: PRP
ORGANISM: E. coli
FEATURE:
LOCATION: 45..65
OTHER INFORMATION: isolated or synthetic EtxB beta4-alpha2 loop fragment derivable
US-09-786-648-4

Query Match
Best Local Similarity 94.4%; Score 101; DB 21; Length 21;
Pred. No. 1.4e-09;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSOHIDSQKAI 21
|||||
DB 1 GATFOVEVPGSOHIDSQKAI 21

RESULT 9
US-08-393-334-2
Sequence 2, Application US/08393334
GENERAL INFORMATION:
APPLICANT: Loomore, Sheena M.
APPLICANT: Yacoub, Reza K.
APPLICANT: Zealey, Gavin R.
APPLICANT: Kleln, Michel H.
TITLE OF INVENTION: Expression of Gene Products From
Geometrically Manipulated Strains of Bordetella
NUMBER OF SEQUENCES: 36

CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 330 University Avenue, Suite 701
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/393,334
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24973
REFERENCE/DOCKET NUMBER: 1038-417
TELECOMMUNICATION INFORMATION:
TELEPHONE: 416-595-1155
TELEFAX: 416-595-1163
TELEX: 065-24567 Simbas
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-393-334-2

Query Match
Best Local Similarity 94.4%; Score 101; DB 7; Length 103;
Pred. No. 1.1e-08;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSOHIDSQKAI 21
|||||
DB 45 GATFOVEVPGSOHIDSQKAI 65

RESULT 10
US-08-782-832-15
Sequence 15, Application US/08782832
GENERAL INFORMATION:
APPLICANT: Arntzen, Charles J.
APPLICANT: Mason, Hugh S.
TITLE OF INVENTION: PRODUCTION OF AN ORALLY IMMUNOGENIC
BACTERIAL PROTEIN IN TRANSGENIC PLANTS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pravel, Hewitt, Kimball & Krieger
STREET: 1177 West Loop South, 10th Floor
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77027-9095
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/782,832
FILING DATE: 13-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/328,716
FILING DATE: 24-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Kettelberger, Denise M.

ORGANISM: synthetic construct
US-09-836-433-22

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 09:04:12 ; Search time 46.52 Seconds

(Without alignments)
45.673 Million cell updates/sec

Title: US-09-786-648-5

Sequence: 1 GETFOVEYPSQHSOKKAI 21

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 308740 seqs, 101176262 residues

Total number of hits satisfying chosen parameters: 308740

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Maximum Match 1008
Listing first 45 summaries

Database : Pending Patents, AA, New.*

1: /cgn2_6/prodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/prodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/prodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/prodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/prodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/prodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/prodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	100.0	103	6	US-10-110-364-22
2	107	100.0	123	6	US-10-110-364-20
3	107	100.0	124	6	US-10-110-364-16
4	107	100.0	124	6	US-10-110-364-18
5	107	100.0	124	6	US-10-110-364-23
6	101	94.4	103	6	US-10-110-364-10
7	101	94.4	103	6	US-10-110-364-13
8	101	94.4	105	6	US-10-110-364-12
9	101	94.4	123	6	US-10-110-364-17
10	101	94.4	124	6	US-10-110-364-2
11	101	94.4	124	6	US-10-110-364-6
12	101	94.4	124	6	US-10-110-364-11
13	101	94.4	124	6	US-10-110-364-19
14	101	94.4	124	6	US-10-110-364-21
15	101	94.4	125	6	US-10-110-364-15
16	101	94.4	138	6	US-10-141-627-2
17	95	88.8	103	6	US-10-110-364-8
18	92	86.0	103	6	US-10-110-364-7
19	89	83.2	103	6	US-10-110-364-5
20	89	83.2	124	6	US-10-110-364-4
21	89	83.2	124	6	US-10-110-364-9
22	89	83.2	382	5	US-09-809-033A-3
23	89	83.2	382	5	US-09-809-033A-4
24	45	42.1	84	5	US-09-620-393B-3333
25	45	42.1	138	5	US-09-620-393B-3332
26	45	42.1	211	5	US-09-540-209B-9908

27	45	42.1	549	5	US-09-540-209B-6846	Sequence 6846, Ap
28	44	41.1	371	6	US-10-155-881-29649	Sequence 29649, A
29	44	41.1	890	7	US-60-360-039-3507	Sequence 3507, Ap
30	43.5	40.7	641	5	US-09-573-655B-1493	Sequence 1493, Ap
31	43.5	40.7	641	7	US-60-382-898-321	Sequence 321, App
32	43	40.2	747	7	US-60-360-039-7808	Sequence 7808, Ap
33	43	40.2	1810	4	US-08-793-273C-4	Sequence 4, Appl1
34	42	39.3	369	6	US-10-138-701-36	Sequence 36, Appl
35	42	39.3	492	6	US-10-155-881-18930	Sequence 18930, A
36	42	39.3	782	1	PCT-US02-07826-312	Sequence 312, App
37	42	39.3	782	6	US-10-097-340-312	Sequence 312, App
38	42	39.3	855	1	PCT-US02-09039-2	Sequence 2, Appl1
39	42	39.3	855	5	US-09-654-600A-2	Sequence 2, Appl1
40	42	39.3	855	6	US-10-104-271-2	Sequence 2, Appl1
41	42	39.3	855	6	US-10-099-700A-2	Sequence 2, Appl1
42	42	39.3	1287	5	US-09-573-655B-1370	Sequence 1370, Ap
43	42	39.3	2359	6	US-10-155-881-28520	Sequence 28520, A
44	41.5	38.8	187	6	US-10-155-881-22303	Sequence 22303, A
45	41.5	38.8	209	6	US-10-155-881-33927	Sequence 33927, A

ALIGNMENTS

RESULT 1
US-10-110-364-22
; Sequence 22, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparaanta, Tapio
; APPLICANT: Ewall, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; PRIOR FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(103)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 494265.
US-10-110-364-22

Query Match 100.0%; Score 107; DB 6; Length 103;
Best Local Similarity 100.0%; Pred. No. 6, 3e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 45 GETFOVEYPSQHSOKKAI 21
|||||
45 GETFOVEYPSQHSOKKAI 65

RESULT 2
US-10-110-364-20
; Sequence 20, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparaanta, Tapio
; APPLICANT: Ewall, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364

```
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 123
; TYPE: PRF
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(123)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 223254.
US-10-110-364-20
```

```
Query Match          100.0%; Score 107; DB 6; Length 123;
Best Local Similarity 100.0%; Pred. No. 7.9e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 GETFOVEVPGSOHIDSOKKAI 21
Db 66 GETFOVEVPGSOHIDSOKKAI 85
```

```
RESULT 3
US-10-110-364-16
; Sequence 16, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 124
; TYPE: PRF
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 1169505.
US-10-110-364-16
```

```
Query Match          100.0%; Score 107; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 7.9e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 GETFOVEVPGSOHIDSOKKAI 21
Db 66 GETFOVEVPGSOHIDSOKKAI 86
```

```
RESULT 4
US-10-110-364-18
; Sequence 18, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
```

```
; TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS
; FILE REFERENCE: ACTBIO.004A
; CURRENT FILING DATE: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 124
; TYPE: PRF
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 145833.
US-10-110-364-18
```

```
Query Match          100.0%; Score 107; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 7.9e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 GETFOVEVPGSOHIDSOKKAI 21
Db 66 GETFOVEVPGSOHIDSOKKAI 86
```

```
RESULT 5
US-10-110-364-23
; Sequence 23, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 124
; TYPE: PRF
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 69630.
US-10-110-364-23
```

```
Query Match          100.0%; Score 107; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 7.9e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 GETFOVEVPGSOHIDSOKKAI 21
Db 66 GETFOVEVPGSOHIDSOKKAI 86
```

```
RESULT 6
US-10-110-364-10
; Sequence 10, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
```

```

1  APPLICANT: Haaparanta, Tapio
2  APPLICANT: Ewalt, Karla L.
3  TITLE OF INVENTION: Aβ5 TOXIN B SUBUNIT MUTANTS WITH ALTERED
4  TITLE OF INVENTION: CHEMICAL CONUGATION CHARACTERISTICS
5  FILE REFERENCE: ACTBIO.0004
6  CURRENT APPLICATION NUMBER: US/10/110.364
7  CURRENT FILING DATE: 2002-04-05
8  PRIOR APPLICATION NUMBER: PCT/US00/27607
9  PRIOR FILING DATE: 2000-10-05
10 PRIOR APPLICATION NUMBER: 60/158,561
11 PRIOR FILING DATE: 1999-10-08
12 NUMBER OF SEQ ID NOS: 31
13 SOFTWARE: FastSeq for Windows Version 4.0
14 SEQ ID NO 10
15 LENGTH: 103
16 TYPE: PRT
17 ORGANISM: vibrio cholera
18 FEATURE:
19 NAME/KEY: VARIANT
20 LOCATION: (1)...(103)
21 OTHER INFORMATION: CTB variant from NCBI gene bank GI: 14215111
22 US-10-110-364-10

```

Query Match	94.4%	Score 101;	DB 6;	Length 103;
Best Local Similarity	95.2%;	Pred. No. 6.7e-10;		
Matches 20;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

```
QY      1 GETFQVEVPGSQHIDSQKAI 21
          | | | | | | | | | | | | |
Db      45 GATFQVEVPGSQHIDSQKAI 65
```

```

RESULT 7
US-10-110-364-13
: Sequence 13, Application US/10110364
: GENERAL INFORMATION:
: APPLICANT: Handley, Harold H.
: APPLICANT: Haaparenta, Tapio
: APPLICANT: Ewalt, Karla L.
: TITLE OF INVENTION: A&S TOXIN B SUBUNIT MUTANTS WITH ALTERED
: TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS
: FILE REFERENCE: ACTBIO.004A
: CURRENT APPLICATION NUMBER: US/10-110, 364
: PRIOR FILING DATE: 2002-04-05
: PRIOR APPLICATION NUMBER: PCT/US00/27607
: PRIOR FILING DATE: 2000-10-05
: PRIOR APPLICATION NUMBER: 60/158,561
: PRIOR FILING DATE: 1999-10-08
: NUMBER OF SEQ ID NOS: 31
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 13
: LENGTH: 103
: TYPE: PRT
: ORGANISM: Vibrio cholera
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (1)...(103)
: OTHER INFORMATION: CTB variant from NCBI gene bank GI: 14215235
: OTHER INFORMATION: (ogawa 41 R35D).
US-10-110-364-13

```

Query Match	94.48;	Score 101;	DB 6;	Length 103;
Best Local Similarity	95.28;	Pred. No. 6.7e-10;		
Matches 20;	Conservative	0;	Mismatches 1;	Indels 0;
				Gaps 0;

```
Qy 1 GETFQVEVPGSQHIDSQKKA1 21
    | | | | | | | | | | | | |
Db 45 GATFQVEVPGSQHIDSQKKA1 65
```

RESULT 8

```

US-10-110-364-12
: Sequence 12, Application US/10110364
: GENERAL INFORMATION:
: APPLICANT: Handley, Harold H.
: APPLICANT: Haaparanta, Tapio
: APPLICANT: Ewall, Karla L.
: TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
: TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS
: FILE REFERENCE: ACTBIO.004A
: CURRENT APPLICATION NUMBER: US/10/110,364
: CURRENT FILING DATE: 2002-04-05
: PRIOR APPLICATION NUMBER: PCT/US00/27607
: PRIOR FILING DATE: 2000-10-05
: PRIOR APPLICATION NUMBER: 60/158,561
: PRIOR FILING DATE: 1998-10-08
: NUMBER OF SEQ ID NOS: 31
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 12
: LENGTH: 105
: TYPE: PRM
: ORGANISM: Vibrio cholera
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (1)...(105)
: OTHER INFORMATION: CTB variant from NCBI gene bank GI: 2781121 (ogawa)
: OTHER INFORMATION: 41).
US-10-110-364-12

```

Query Match	94.48;	Score 101;	DB 6;	Length 105;
Best Local Similarity	95.28;	Pred. No. 6.8e-10;		
Matches 20;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0

```
QY      1 GETFOVEVPGSQHIDSQKKAI 21
          | | | | | | | | | | | | |
Db      46 GATFOVEVPGSQHIDSQKKAI 66
```

```

RESULT      9
US-10-110-364-17
Sequence 17, Application US/10110364

GENERAL INFORMATION:
APPLICANT: Handley, Harold H.
APPLICANT: Haaparanta, Tapio
APPLICANT: Ewelt, Karla L.
TITLE OF INVENTION: A85 TOXIN B SUBUNIT MUTANTS WITH ALTERED
TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS
FILE REFERENCE: ACTBIO.004A
CURRENT APPLICATION NUMBER: US/10/110,364
CURRENT FILING DATE: 2002-04-05
PRIOR APPLICATION NUMBER: PCT/US00/27607
PRIOR FILING DATE: 2000-10-05
PRIOR APPLICATION NUMBER: 60/158,561
PRIOR FILING DATE: 1999-10-08
NUMBER OF SEQ ID NOS: 31
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17
LENGTH: 123
TYPE: PRT
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(123)
OTHER INFORMATION: LTB variant from NCBI gene bank GI: 1395122
US-10-110-364-17

```

Query Match	94.48;	Score 101;	DB 6;	Length 123;
Best Local Similarity	95.28;	Pred. No. 8.3e-10;		
Matches 20;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

[illegible]

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GATFOVEVPGSQHIDSOKKAI 21
Db 66 GATFOVEVPGSQHIDSOKKAI 86

RESULT 14

US-10-110-364-21
; Sequence 21, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; PRIOR FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 408996.
US-10-110-364-21

Query Match 94.4%; Score 101; DB 6; Length 124;
Best Local Similarity 95.2%; Pred. No. 8.3e-10;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21
Db 66 GATFOVEVPGSQHIDSOKKAI 86

RESULT 15

US-10-110-364-15
; Sequence 15, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; PRIOR FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 125
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(125)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 3062900.
US-10-110-364-15

Query Match 94.4%; Score 101; DB 6; Length 125;
Best Local Similarity 95.2%; Pred. No. 8.4e-10;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GATFOVEVPGSQHIDSOKKAI 21
Db 67 GATFOVEVPGSQHIDSOKKAI 87

Search completed: July 3, 2002, 09:04:12
Job time: 652 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 08:56:33 ; Search time 46.57 seconds
(without alignments)
43.330 Million cell updates/sec

Title: US-09-786-648-5

Perfect score: 107
Sequence: 1 GETFGVEVPGSGHDSOKKAI 21

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR71:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	124	1 QLECB	heat-labile entero
2	95	88.8	124	1 XVVCB	cholera enterotoxi
3	49	45.8	255	2 A86457	probable peptide c
4	47	43.9	392	2 T04150	RAD23 protein homo
5	46	43.0	91	2 C96580	hypothetical prote
6	46	43.0	574	2 C86400	hypothetical prote
7	45	42.1	439	2 T49189	kinesin heavy chain
8	45	42.1	2733	2 S15760	genome polypeptide
9	44	41.1	263	2 S78364	conserved hypochet
10	44	41.1	374	2 T19866	hypothetical prote
11	44	41.1	864	2 T49574	probable carnitine
12	44	41.1	2731	1 VFIHJH	genome polypeptide
13	43.5	40.7	641	2 C84726	probable receptor
14	43	40.2	128	1 UODOR	ubiquitin / riboso
15	43	40.2	154	1 UODOR7	ubiquitin / riboso
16	43	40.2	159	2 A97466	hypothetical prote
17	43	40.2	159	2 AC2684	hypothetical prote
18	43	40.2	228	2 D34080	ubiquitin 18 - sll
19	43	40.2	229	2 B27806	ubiquitin (clone 1
20	43	40.2	368	2 T04861	hypothetical prote
21	43	40.2	380	2 C34080	polyubiquitin 5 (c
22	43	40.2	380	2 B34080	polyubiquitin 5 (c
23	43	40.2	381	2 A27806	polyubiquitin 5 (c
24	43	40.2	532	2 A34080	polyubiquitin 7 (c
25	43	40.2	1810	1 A32230	tenascin precursor
26	42	39.3	204	2 E82963	probable peptide c
27	42	39.3	330	2 C89848	peptide chain rele
28	42	39.3	427	2 JC5694	stress-activated p
29	42	39.3	672	2 A87441	penicillin-binding

30	42	39.3	1742	2 S76110	hypothetical prote
31	42	39.3	1959	2 AG1085	hypothetical prote
32	41.5	38.8	376	1 S17246	chorismate synthas
33	41.5	38.8	500	2 UC4022	4-aminobutyrate tr
34	41	38.3	166	2 G90661	probable peptide c
35	41	38.3	166	2 E64748	translation releas
36	41	38.3	166	2 E85512	probable peptide c
37	41	38.3	204	2 A10542	probable peptide c
38	41	38.3	260	2 C96827	protein F20B17.2 (
39	41	38.3	269	1 A25973	pertussis toxin ch
40	41	38.3	313	2 E88216	protein B0495.8 (1
41	41	38.3	333	2 D87512	alcohol dehydrogen
42	41	38.3	386	1 S66056	yeast protein - Bac
43	41	38.3	399	2 T46898	queuine tRNA-ribos
44	41	38.3	550	2 T01770	hypothetical prote
45	41	38.3	683	2 S01433	repressor protein

ALIGNMENTS

RESULT 1
QLECB
heat-labile enterotoxin chain B precursor - Escherichia coli
C:Species: Escherichia coli
C>Date: 29-Jun-1981 #sequence: revision 29-Jun-1981 #text: change 18-Jun-1999
C:Accession: A01820; B26946; I41194; I41287; I67644; A61475
R:Dallas, W.S.; Falkow, S.
Nature 288, 499-501, 1980
A:Title: Amino acid sequence homology between cholera toxin and Escherichia coli heat
A:Reference number: A01820; MVID:81074965
A:Accession: A01820
A:Molecule type: mRNA
A:Residues: 1-124 <DAL>
R:Yamamoto, T.; Gojobori, T.; Yokota, T.
J. Bacteriol. 169, 1352-1357, 1987
A:Title: Evolutionary origin of pathogenic determinants in enterotoxigenic Escherich
A:Reference number: A26946; MVID:87137303
A:Accession: B26946
A:Molecule type: DNA
A:Residues: 1-27, 'E', '29-63, 'K', '65-124 <YAM>
A:Reference number: I41194; MVID:8516481
A:Accession: I41194
A:Molecule type: DNA
A:Residues: 1-5, 'F', '7-17, 'C', '19-24, 'S', '26-27, 'E', '29-33, 'H', '35-63, 'K', '65-66, 'A', '68-122
A:Cross-references: GB:M17874; NID:g145830; PIDN:AAA98064.1; PID:g145831
A:Experimental source: Plasmid ENT-R PCG86
R:Ibrahim, I.; Gentz, R.
J. Biol. Chem. 262, 10189-10194, 1987
A:Title: A functional interaction between the signal peptide and the translation appa
Ciculum.
A:Reference number: I41287; MVID:87280041
A:Accession: I41287
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-22 <RE2>
A:Cross-references: GB:M17101; NID:g146375; PIDN:AAA23973.1; PID:g146376
R:Imoue, T.; Tsuji, T.; Koto, M.; Imamura, S.; Miyama, A.
FEBS Microbiol. Lett. 108, 157-161, 1993
A:Title: Amino acid sequence of heat-labile enterotoxin from chicken enterotoxigenic
A:Reference number: I53542; MVID:93252225
A:Accession: I67644
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-17, 'C', '19, 'V', '21-24, 'S', '26-27, 'E', '29-63, 'K', '65-66, 'A', '68-122, 'E', '124 <R
A:Cross-references: GB:S60731; NID:g408994; PIDN:AAC60441.1; PID:g408996
R:Tsuji, T.; Iida, T.; Honda, T.; Mawatani, T.; Nagahama, M.; Sakurai, J.; Wada, K.;
Microb. Pathog. 2, 381-390, 1987

A:Title: A unique amino acid sequence of the B subunit of a heat-labile enterotoxin isol
A:Reference number: A61475; MUID:89180953
A:Accession: A61475
A:Molecule type: protein
A:Residues: 22-24,'S',26-27,'E',29-63,'K',65-66,'A',68-95,'A',97-122,'E',124 <TSU>
A:Experimental source: strain 240-3
C:Function: the heat-labile enterotoxin molecule contains one A chain and five or six B
C:Complex: the heat-labile enterotoxin molecule contains one A chain and five or six B
A:Description: the biological activity of the toxin is produced by the A chain, which ac
C:Superfamily: cholera enterotoxin beta chain
C:Keywords: enterotoxin
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-124/Product: heat-labile enterotoxin chain B #status predicted <MAT>
F:30-107/Disulfide bonds: #status predicted

Query Match 100.0%; Score 107; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 5.9e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSHIDSOKKAI 21
|||||
DB 66 GETFOVEVPGSHIDSOKKAI 86

RESULT 2
XVVCB
cholera enterotoxin, B chain precursor VC1456 [validated] - Vibrio cholerae (strain N169
N:Alternate names: enterotoxin beta chain
C:Species: Vibrio cholerae
C:Date: 24-Apr-1984 #sequence_revision 01-Sep-2000 #text_change 02-Feb-2001
C:Accession: S14624; S39238; S39241; H82196; JCI078; S17666; PC1010; A05130; A01819; A38
R:Dams, E.; de Wolf, M.; Dierick, W.
submitted to the EMBL Data Library, March 1991
A:Description: Correction of the cholera toxin nucleotide sequence of the Vibrio cholera
A:Reference number: S14623
A:Accession: S14624
A:Molecule type: DNA
A:Residues: 1-124 <DNA>
A:Cross-references: EMBL:X58786; NID:948420; PIDN:CAA41593.1; PID:948422
A:Experimental source: strain 2125
R:Lebens, M.; Holmgren, J.
submitted to the EMBL Data Library, November 1993
A:Description: Structure and arrangement of the Cholera toxin genes in Vibrio Cholera 01
A:Reference number: S39238
A:Accession: S39238
A:Molecule type: DNA
A:Residues: 1-124 <LEB>
A:Cross-references: EMBL:X76390; NID:9433856; PIDN:CAA53973.1; PID:9433857
A:Accession: S39241
A:Molecule type: DNA
A:Residues: 1-124 <LEW>
A:Cross-references: EMBL:X76391; NID:9433859; PIDN:CAA53976.1; PID:9433861
R:Heideberg, J.F.; Elsen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Charidson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoti, I.; Sellers, F
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: H82196
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-124 <HEI>
A:Cross-references: GB:AE004224; GB:AE003852; NID:99655952; PIDN:AAF94613.1; GSPDS:GN001
R:Shi, C.H.; Cao, C.; Zhang, J.S.; Ma, Q.J.
Chinese Biochem. J. 9, 395-399, 1993
A:Title: Nucleotide sequence analysis of the gene encoding the classical biotype V.chol
A:Reference number: JCI078
A:Accession: JCI078
A:Molecule type: DNA
A:Residues: 1-20,'Q',22-31,'Q',33-38,'H',40-49,'G',51-67,'T',69-124 <SHI>
A:Experimental source: classical biotype strain 5698

R:Dams, E.; de Wolf, M.; Dierick, W.
Biochim. Biophys. Acta 1090, 139-141, 1991
A:Title: Nucleotide sequence analysis of the CT operon of the Vibrio cholerae classic
A:Reference number: S17665; MUID:91355224
A:Accession: S17666
A:Molecule type: DNA
A:Residues: 1-38,'H',40-67,'T',69-124 <DA2>
A:Cross-references: EMBL:X58785; NID:948888; PIDN:CAA41591.1; PID:948890
R:Ma, Q.J.; Liu, C.X.; Xiong, L.S.; Yu, X.Q.
Sci. Sin. B Chem. Biol. Agric. Med. Earth Sci. 34, 274-280, 1991
A:Title: B subunit of cholera toxin produced in Escherichia coli.
A:Reference number: PC1010
A:Accession: PC1010
A:Molecule type: protein
A:Residues: 22-38,'H',40-41 <MAQ>
R:Mekalanos, J.J.; Swartz, D.J.; Pearson, G.D.N.; Harford, N.; Groyne, F.; de Wilde,
Nature 306, 551-557, 1983
A:Reference number: A93320; MUID:84068199
A:Accession: A05130
A:Molecule type: DNA
A:Residues: 1-32,'S',34-74,'S',76-124 <MEK>
A:Cross-references: GB:X00171; NID:948347; PIDN:CAA24996.1; PID:9758351
R:Kurosky, A.; Markel, D.E.; Peterson, J.W.
J. Biol. Chem. 252, 7257-7264, 1977
A:Title: Covalent structure of the beta chain of cholera enterotoxin.
A:Reference number: A01819; MUID:78005537
A:Accession: A01819
A:Molecule type: protein
A:Residues: 22-38,'H',40-42,'N',44-67,'T',69-90,'N',92-124 <KUR>
R:Lai, C.Y.
J. Biol. Chem. 252, 7249-7256, 1977
A:Title: Determination of the primary structure of cholera toxin B subunit.
A:Reference number: A38033; MUID:78005536
A:Accession: A38033
A:Molecule type: protein
A:Residues: 22-38,'H',40-42,'N',44-67,'T',69,'E',71,'Q',74-75,'VE',78-86,'Q',88-99,'Q',101-103
A:Note: the difference at residue 70 may be due to deamidation during preparation
R:Nakashima, Y.; Napierkowski, P.; Schafer, D.E.; Konigsberg, W.H.
FEBS Lett. 68, 275-278, 1976
A:Title: Primary structure of the B subunit of cholera enterotoxin.
A:Reference number: A38034; MUID:77026365
A:Accession: A38034
A:Molecule type: protein
A:Residues: 22-38,'H',40-67,'T',69,'E',71,'Q',74-75,'VE',78-86,'Q',88-99,'Q',101-103
R:Takao, T.; Matanabe, H.; Shimomishi, Y.
Eur. J. Biochem. 146, 503-508, 1985
A:Title: Facile identification of protein sequences by mass spectrometry.
A:Reference number: A21910; MUID:85126976
A:Accession: A21910
A:Molecule type: protein
A:Residues: 22-38,'H',40-42,'N',44-67,'T',69-90,'N',92-124 <TAK>
A:Experimental source: biotype Inaba 569B
A:Note: Asn-65 was partially deamidated to Asp
C:Comment: The authors translated the codon TCA for residue 33 as Tyr.
C:Genetics:
A:Gene: VC1456
A:Map position: 1
C:Complex: the cholera enterotoxin molecule contains three kinds of chains; an alpha
cate noncovalently with the subunit B, an aggregate of five beta chains
C:Function:
A:Description: Involved in binding of the toxin to cell membranes
C:Superfamily: cholera enterotoxin beta chain
C:Keywords: enterotoxin; toxin
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-124/Product: cholera enterotoxin chain B #status predicted <MAT>
F:30-107/Disulfide bonds: #status experimental

Query Match 88.8%; Score 95; DB 1; Length 124;
Best Local Similarity 90.5%; Pred. No. 5.2e-08;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSHIDSOKKAI 21

Db 66 GAIFOVEVPGSQHIDSOKKAI 86

RESULT 3

A86457

C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Nov-2001

C:Accession: A86457

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzila

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: A86457

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-255 <STO>

A:Cross-references: GB:AE005172; NID:g10645493; PIDN:AAG21605.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 45.8%; Score 49; DB 2; Length 255;

Best Local Similarity 45.0%; Pred. No. 3.1;

Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 2 ETFOVEVPGSQHIDSOKKAI 21

||||| |||||

Db 101 ETFRVSGCGCHRNKRDSAV 120

RESULT 4

T04150

RAD23 protein homolog - rice

C:Species: Oryza sativa (rice)

C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000

C:Accession: T04150

R:Schultz, T.F.; Quatrano, R.S.

Plant Mol. Biol. 34, 557-562, 1997

A>Title: Characterization and expression of a rice RAD23 gene.

A:Reference number: Z08695; MUID:97369378

A:Accession: T04150

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-392 <SCCH>

A:Cross-references: EMBL:U63530; NID:g1488296; PIDN:AAB65841.1; PID:g1488297

A:Experimental source: cv. Nipponbare

C:Genetics:

A:Gene: RAD23

C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match 43.9%; Score 47; DB 2; Length 392;

Best Local Similarity 42.9%; Pred. No. 10;

Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

OY 1 GTFQVEVPGSQHIDSOKKAI 21

||||| :| :|

Db 10 GSTFOIEVDSQKQVADVKRII 30

RESULT 5

C96580

hypothetical protein F1511.6 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: C96580

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzila

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: C96580

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-91 <STO>

A:Cross-references: GB:AE005173; NID:g4587538; PIDN:AAD25769.1; GSPDB:GN00141

C:Genetics:

A:Gene: F1511.6

A:Map position: 1

Query Match 43.0%; Score 46; DB 2; Length 91;

Best Local Similarity 47.6%; Pred. No. 3.2;

Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 1 GTFQVEVPGSQHIDSOKKAI 21

||||| |||||

Db 58 GKTFLVKGSEITQVKNMI 78

RESULT 6

C86400

hypothetical protein T17H3.2 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001

C:Accession: C86400

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzila

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: C86400

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-574 <STO>

A:Cross-references: GB:AE005172; NID:g5668763; PIDN:AAD45990.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 43.0%; Score 46; DB 2; Length 574;

Best Local Similarity 58.8%; Pred. No. 23;

Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 TFQVEVPGSQHIDSOKK 19

||||| |||||

Db 19 TFFVVDPSQHIEVKK 35

RESULT 7

T49189

kinesin heavy chain-like protein - Arabidopsis thaliana

N:Alternate names: protein MAA21.110

C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000

Db 541 QTFWADAPGDAKIDALRKA 559

RESULT 12

VF1JH

genome polypeptide 1b - murine hepatitis virus (strain JHM)

N:Alternate names: 1b protein

C:Species: murine hepatitis virus, MHV

C>Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 11-Jun-1999

C:Accession: B36815

R:Lee, H.J.; Shieh, C.K.; Gorbelenya, A.E.; Koonin, E.V.; La Monica, N.; Tuler, J.; Bagc submitted to Genbank, February 1991

A:Description: The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding

A:Reference number: A36815

A:Accession: B36815

A:Molecule type: genomic RNA

A:Residues: 1-2731 <LEE>

A:Cross-references: GB:M55148; NID:g331851; PIDN:AAA46458.1; PID:g331853

R:Lee, H.J.; Shieh, C.K.; Gorbelenya, A.E.; Koonin, E.V.; La Monica, N.; Tuler, J.; Bagc Virology 180, 567-582, 1991

A:Title: The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the

A:Reference number: A38547; MUID:91111976

A:Contents: annotation

A>Note: neither nucleotide nor complete amino acid sequence is given

C:Comment: This protein may be translated as a 1a-1b polyprotein by a ribosomal frameshift

C:Genetics:

A:Gene: 1b

C:Superfamily: Infectious bronchitis virus RNA-directed RNA polymerase

C:Keywords: glycoprotein; nucleotidyltransferase; RNA biosynthesis

F:269,304,785,1164,1287,1524,1842,2196,2575,2630,2645,2665/Binding site: carbohydrate (A

Query Match 41.1%; Score 44; DB 1; Length 2731;

Best Local Similarity 69.2%; Pred. No. 2.5e+02;

Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ETFQVEVPGSOHI 14

Db 1198 ETFQNNVNYOHI 1210

RESULT 13

C84726

Probable receptor-like protein kinase [Imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: C84726

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.

Neuss, D.; Nierman, W.C.; White, O.; Eissen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487

A:Accession: C84726

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-641 <STO>

A:Cross-references: GB:AE002093; NID:g4887748; PIDN:AAD32284.1; GSPDB:GN00139

C:Genetics:

A:Gene: AT2G31880

A:Map position: 2

Query Match

Best Local Similarity 40.7%; Score 43.5; DB 2; Length 641;

Matches 11; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

OY 1 GETFOVEVPGSO-HIDSOKKAI 21

Db 359 GEVFKAEPLSGSKIIAVKKVI 380

RESULT 14

UQDOR

ubiquitin / ribosomal protein CEP52 - slime mold (Dictyostelium discoideum)

N:Alternate names: ubiquitin fusion protein

C:Species: Dictyostelium discoideum

C>Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 22-Jun-1999

C:Accession: S00357; A25863

R:Muller-Traubnerberger, A.; Westphal, M.; Jaeger, E.; Noegel, A.; Gerisch, G.

FEBS Lett. 229, 273-278, 1988

A:Title: Complete cDNA sequence of a Dictyostelium ubiquitin with a carboxy-terminal

A:Reference number: S00357; MUID:88152253

A:Accession: S00357

A:Molecule type: mRNA

A:Residues: 1-128 <ME>

A:Cross-references: EMBL:X07210; NID:g7381; PIDN:CAA30183.1; PID:g7382

A:Experimental source: strain AX2-214

C:Genetics:

A:Gene: DUB1

C:Superfamily: ubiquitin/ribosomal protein CEP52; ribosomal protein CEP52 homology; u

C:Keywords: DNA binding; protein biosynthesis; protein degradation; ribosome; zinc fi

F:1-76/Product: ubiquitin #status predicted <UBI>

F:77-128/Product: ribosomal protein CEP52 #status predicted <RIB>

F:95-114/Region: zinc finger CCCC motif

F:121-128/Region: nuclear location signal

Query Match 40.2%; Score 43; DB 1; Length 128;

Best Local Similarity 42.9%; Pred. No. 14;

Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSOHIHDSOKKAI 21

Db 10 GKTITLVEGSDNIENYKAKI 30

RESULT 15

UQDOR7

ubiquitin / ribosomal protein S27a - slime mold (Dictyostelium discoideum)

C:Species: Dictyostelium discoideum

C>Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 21-Jul-2000

C:Accession: E34080

R:Omachi, T.; Giorda, R.; Shaw, D.R.; Ennis, H.L.

Biochemistry 28, 5226-5231, 1989

A:Title: Molecular organization of developmentally regulated Dictyostelium discoideum

A:Reference number: A34080; MUID:89352609

A:Accession: E34080

A:Molecule type: mRNA

A:Residues: 1-154 <OHM>

A:Cross-references: GB:J02858; NID:g167940; PIDN:AAA33264.1; PID:g167941

C:Superfamily: ubiquitin / rat ribosomal protein S27a; ribosomal protein S27a homology

C:Keywords: protein biosynthesis; protein degradation; ribosome

F:1-76/Product: ubiquitin #status predicted <MAT1>

F:1-76/Domain: ubiquitin homology <UBH>

F:77-154/Product: ribosomal protein S27a #status predicted <MAT2>

F:102-152/Domain: ribosomal protein S27a homology <RIB>

Query Match

Best Local Similarity 40.2%; Score 43; DB 1; Length 154;

Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSOHIHDSOKKAI 21

Db 10 GKTITLVEGSDNIENYKAKI 30

Search completed: July 3, 2002, 08:56:34
Job time: 234 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 09:04:43 ; Search time 21.51 Seconds

(without alignments)
37.802 Million cell updates/sec

Title: US-09-786-648-5

Sequence: 1 GETFOVEYPSQSHIDSOKKAI 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	100.0	124	1	P32890 escherichia
2	101	94.4	124	1	ELBP_ECOLI
3	95	88.8	124	1	CHMB_VIBCH
4	45	42.1	273	1	RRPB_CVMA5
5	44	41.1	263	1	YC43_ODOSI
6	44	41.1	467	1	IRF6_HUMAN
7	44	41.1	467	1	IRF6_MOUSE
8	44	41.1	2731	1	RRPB_CVMA5
9	43	40.2	76	1	UBIO_DICDI
10	43	40.2	1808	1	TENA_CHICK
11	42	39.3	855	1	ST14_HUMAN
12	41.5	38.8	376	1	AROC_YEAST
13	41.5	38.8	500	1	GABP_HUMAN
14	41	38.3	141	1	RFB_ECOLI
15	41	38.3	313	1	YP68_CAEEL
16	41	38.3	385	1	TG7_ZYMO
17	41	38.3	386	1	YAAN_BACSU
18	41	38.3	567	1	STP_STRCO
19	41	38.3	683	1	RPC_BPPHC
20	41	38.3	1702	1	DPOL_THEHI
21	41	38.3	1829	1	DPOL_THEST
22	40.5	37.9	862	1	PMS2_HUMAN
23	40	37.4	76	1	UBIO_NEUCR
24	40	37.4	148	1	IF5A_AERPE
25	40	37.4	308	1	HS74_CAEEL
26	40	37.4	289	1	CLT2_RAT
27	40	37.4	355	1	YRY1_CAEEL
28	40	37.4	395	1	DBH3_ENTHI
29	40	37.4	523	1	DBP3_YEAST
30	40	37.4	773	1	DPOL_THEGO
31	40	37.4	775	1	DPOL_THES9
32	40	37.4	1048	1	CPXB_BACME
33	40	37.4	1523	1	DPOL_THIEFM

ALIGNMENTS

RESULT	ID	ELBP_ECOLI	STANDARD:	PRT:	124 AA.
AC	P32890	ELBP_ECOLI	P01557		
DT	21-JUL-1986	(Rel. 01, Created)			
DT	01-NOV-1995	(Rel. 32, Last sequence update)			
DT	15-JUL-1998	(Rel. 36, Last annotation update)			
DE	Heat-labile enterotoxin B chain precursor (LT-B, porcine) (LTP-B).				
GN	ELTB OR LTPB.				
OS	Escherichia coli.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OC	Escherichia.				
OX	NCBI_TaxID=562;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-ISOLATE P307;				
RX	MEDLINE=81074965; PubMed=7003397;				
RA	Dallas W.S., Falkow S.;				
RT	"Amino acid sequence homology between cholera toxin and Escherichia				
RT	coli heat-labile toxin.";				
RL	Nature 288:499-501(1980).				
RN	[2]				
RP	REVISIONS TO 28 AND 64.				
RC	STRAIN-ISOLATE P307;				
RX	MEDLINE=85156481; PubMed=3884513;				
RA	Jeong J., Vinal A.C., Dallas W.S.;				
RT	"Evolutionary origin of pathogenic determinants in enterotoxigenic				
RT	Escherichia coli and Vibrio cholerae O1.";				
RL	J. Bacteriol. 169:1352-1357(1987).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-ISOLATE PCG86;				
RX	MEDLINE=87137303; PubMed=3546273;				
RA	Yamamoto T., Gojobori T., Yokota T.;				
RT	"A functional interaction between the signal peptide and the				
RT	translation apparatus is detected by the use of a single point				
RT	mutation which blocks translocation across mammalian endoplasmic				
RT	reticulum.";				
RL	J. Biol. Chem. 262:10189-10194(1987).				
RN	[5]				
RP	X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).				
RX	MEDLINE=93240541; PubMed=8478941;				
RA	Sixma T.K., van Zanten B.A.M., Dauter Z., Hol W.G.J.;				
RT	"Refined structure of Escherichia coli heat-labile enterotoxin, a				
RT	close relative of cholera toxin.";				
RL	J. Mol. Biol. 230:890-918(1993).				
RN	[6]				
RP	X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).				
RX	MEDLINE=91238966; PubMed=2034287;				

09h05 thermococcus
P77933 pyrococcus
Q9h84 thermococcus
P07206 klebsiella
P04838 saccharomyc
P05820 escherichia
P75049 mycoplasma
Q75570 homo sapien
Q9h840 thermoplasm
P16872 mus musculu
P51882 rattus norv
Q28589 ovis aries

RA Sijma T.K., Pronk S.E., Kalk K.H., Wartna E.S., van Zanten B.A.M.,
 RA Witchoit B., Hol W.G.J.: "Crystal structure of a cholera toxin-related heat-labile enterotoxin
 RT from *E. coli*.";
 RL Nature 351:371-377(1991).
 RN [7]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE=95349400; PubMed=7623669;
 RA Domenighini M., Pizza M., Jobling M.G., Holmes R.K., Rappuoli R.:
 RT "Identification of errors among database sequence entries and
 RT comparison of correct amino acid sequences for the heat-labile
 RT enterotoxin of *Escherichia coli* and *Vibrio cholerae*.";
 RL Mol. Microbiol. 15:1165-1167(1995).
 CC -i- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
 CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
 CC -i- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL, M17873; AAA98065.1; -;
 DR EMBL, M15363; AAA24792.1; -;
 DR EMBL, M17101; AAA23973.1; -;
 DR PIR: A01820; OLECB.
 DR PIR: B26946; OLECB.
 DR PDB: 1LTA; 31-JAN-94.
 DR PDB: 1LTB; 31-JAN-94.
 DR PDB: 1LTG; 15-SEP-95.
 DR PDB: 1LTI; 17-AUG-96.
 DR PDB: 1LTS; 31-JAN-94.
 DR PDB: 1LTT; 07-JUL-97.
 DR PDB: 1LT4; 16-JUN-97.
 DR PDB: 1LT5; 03-DEC-97.
 DR PDB: 1LT6; 03-DEC-97.
 DR PDB: 1LTL; 20-APR-95.
 DR InterPro: IPR001835; Enterotoxin_B.
 DR Pfam: PF01376; Enterotoxin_B.1.
 DR PRINTS: PR00772; ENTEROTOXINB.
 DR ProDom: PD012805; Enterotoxin_B.1.
 KW Enterotoxin; Signal; 3D-structure.
 FT SIGNAL 1 21
 FT CHAIN 22 124 HEAT-LABILE ENTEROTOXIN B CHAIN.
 FT DISULFID 30 107
 FT HELIX 26 30
 FT TURN 31 32
 FT STRAND 36 43
 FT STRAND 47 51
 FT TURN 54 55
 FT STRAND 58 62
 FT TURN 64 65
 FT STRAND 68 71
 FT TURN 76 77
 FT HELIX 80 98
 FT TURN 99 100
 FT STRAND 103 109
 FT STRAND 115 123
 SO SEQUENCE 124 AA; 14133 MW; 6DB7DE58395EA70D CRC64;

Query Match 100.0%; Score 107; DB 1; Length 124;
 Best Local Similarity 100.0%; Pred. No. 3e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GETFOVEVPGSOHDSOKKAI 21
 ||||||||||||||||
 Db 66 GETFOVEVPGSOHDSOKKAI 86

RESULT 2
 EBLH_ECOLI STANDARD; PRT; 124 AA.
 ID P13811;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Heat-labile enterotoxin B chain precursor (LT-B, human) (LTH-B).
 GN ELPB OR LTPB.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Escherichia*.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ISOLATE H74-114;
 RX MEDLINE=85156481; PubMed=3884513;
 RA Leong J., Vinal A.C., Dallas W.S.:
 RT "Nucleotide sequence comparison between heat-labile toxin B-subunit
 RT cistrons from *Escherichia coli* of human and porcine origin.";
 RL Infect. Immun. 48:73-77(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ISOLATE H10407;
 RX MEDLINE=83114628; PubMed=6759877;
 RA Yamamoto T., Tamura T.A., Yokota T., Takano T.:
 RT "Overlapping genes in the heat-labile enterotoxin operon originating
 RT from *Escherichia coli* human strain.";
 RL Mol. Gen. Genet. 188:356-359(1982).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ISOLATE H10407;
 RX MEDLINE=93252225; PubMed=8486242;
 RA Inoue T., Tsuji T., Koto M., Imanura S., Miyama A.:
 RT "Amino acid sequence of heat-labile enterotoxin from chicken
 RT enterotoxigenic *Escherichia coli* is identical to that of human strain
 RT H 10407.";
 RL FEMS Microbiol. Lett. 108:157-161(1993).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ETEC LT 87;
 RA Germain Y., Desperrier J.M.:
 RL Submitted (Oct-1996) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE=95349400; PubMed=7623669;
 RA Domenighini M., Pizza M., Jobling M.G., Holmes R.K., Rappuoli R.:
 RT "Identification of errors among database sequence entries and
 RT comparison of correct amino acid sequences for the heat-labile
 RT enterotoxins of *Escherichia coli* and *Vibrio cholerae*.";
 RL Mol. Microbiol. 15:1165-1167(1995).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (3.04 ANGSTROMS).
 RX MEDLINE=99185101; PubMed=10085117;
 RA Matkovic-Calogovic D., Loregian A., D'Acunzio M.R., Battistutta R.,
 RA Tossi A., Palu G., Zanotti G.:
 RT "Crystal structure of the B subunit of *Escherichia coli* heat-labile
 RT enterotoxin carrying peptides with anti-herpes simplex virus type 1
 RT activity.";
 RL J. Biol. Chem. 274:8764-8769(1999).
 CC -i- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
 CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
 CC -i- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

DR EMBL: M17874; AAA98064.1; -
 DR EMBL: J01646; AAB02982.1; -
 DR EMBL: S60731; AAG00441.1; -
 DR EMBL: X83966; CAA58800.1; -
 DR PDB: 1L7R; 23-MAR-99.
 DR InterPro: IPR001835; Enterotoxin_B.
 DR Pfam: PF01376; Enterotoxin_B; 1.
 DR PRINTS: PR00772; ENTEROTOXINB.
 DR ProDom: PD012805; Enterotoxin_B; 1.
 DR Enterotoxin; Signal; 3D-structure.
 FT SIGNAL 1 21
 FT CHAIN 22 124 HEAT-LABILE ENTEROTOXIN B CHAIN.
 FT DISULFID 30 107
 FT VARIANT 6 6 F -> C (IN ISOLATE H10407).
 FT VARIANT 20 20 H -> Y (IN ISOLATE H10407).
 FT VARIANT 34 34 H -> R (IN ISOLATE H10407).
 FT SEQUENCE 124 AA; 14027 MW; E9F7FC7B9D3BC47 CRC64;
 SQ
 Query Match 94.4%; Score 101; DB 1; Length 124;
 Best Local Similarity 95.2%; Pred. No. 2.7e-09;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 1 GEFQVEVPGSOHIDSKKAI 21
 DB 66 GATFOVEVPGSOHIDSKKAI 86
 RESULT 3
 CHTB_VIBCH STANDARD; PRT; 124 AA.
 AC P01556; Q9J002; (Created)
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cholera enterotoxin, beta chain precursor.
 GN CTXB OR TOXB OR VC1456.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 OX NCBI_TaxID=66;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84061784; PubMed=6315707;
 RA Lockman H., Kaper J.B.;
 RT "Nucleotide sequence analysis of the A2 and B subunits of Vibrio
 cholerae enterotoxin.";
 RL J. Biol. Chem. 258:13722-13726(1983).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-EL TOR 2125;
 RX MEDLINE=84068199; PubMed=6646234;
 RA Mekalanos J.J., Swartz D.J., Pearson G.D.N., Harford N., Groyne F.,
 de Wilde M.;
 RT "Cholera toxin genes: nucleotide sequence, deletion analysis and
 vaccine development.";
 RL Nature 306:551-557(1983).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-EL TOR 2125;
 RA Dams E., de Wolf M., Dierick W.;
 RL Submitted (May-1991) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=4260B / SEROTYPE O139;
 RX MEDLINE=94237453; PubMed=8181723;
 RA Lebens M., Holmgren J.;
 RT "Structure and arrangement of the cholera toxin genes in Vibrio
 cholerae O139.";
 RL FEBS Microbiol. Lett. 117:197-202(1994).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1854 / O139-BENGAL;
 RA Yamamoto K., Do V.G.R.F., Xu M., Iida T., Miwatani T., Albert M.J.,

RA Honda T.;
 RL Submitted (May-1994) to the EMBL/Genbank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN-EL TOR N16961 / SEROTYPE O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 McDonald L., Utterback T., Fleischmann R.D., Nielsen W.C., White O.,
 Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 cholerae.";
 RL Nature 406:477-483(2000).
 RN [7]
 RP SEQUENCE OF 22-124.
 RX MEDLINE=78005537; PubMed=903363;
 RA Kurosky A., Markel D.E., Peterson J.W.;
 RT "Covalent structure of the beta chain of cholera enterotoxin.";
 RL J. Biol. Chem. 252:7257-7264(1977).
 RN [8]
 RP SEQUENCE OF 22-124.
 RX MEDLINE=78005536; PubMed=903362;
 RA Lai C.-Y.;
 RT "Determination of the primary structure of cholera toxin B subunit.";
 RL J. Biol. Chem. 252:7249-7256(1977).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RX MEDLINE=94272319; PubMed=8003954;
 RA Merritt E.A., Sarfaty S., van den Akker F., L'Hoir C., Martini J.A.,
 Hol W.G.J.;
 RT "Crystal structure of cholera toxin B-pentamer bound to receptor GM1
 pentasaccharide.";
 RL Protein Sci. 3:166-175(1994).
 RN [10]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
 RX MEDLINE=95387394; PubMed=7658472;
 RA Zhang R.G., Westbrook M.L., Westbrook E.M., Scott D.L., Olwinowski Z.,
 Maulik P.R., Reed R.A., Shipley G.G.;
 RT "The 2.4 A crystal structure of cholera toxin B subunit pentamer:
 choleraenoid.";
 RL J. Mol. Biol. 251:550-562(1995).
 RN [11]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RC STRAIN-OGAWA 41 / CLASSICAL BIOTYPE;
 RX MEDLINE=97376625; PubMed=9232653;
 RA Merritt E.A., Sarfaty S., Jobling M.G., Chang T., Holmes R.K.,
 Hirst T.R., Hol W.G.J.;
 RT "Structural studies of receptor binding by cholera toxin mutants.";
 RL Protein Sci. 6:1516-1528(1997).
 CC -1- FUNCTION: THE BETA CHAIN AGGREGATE (B SUBUNIT) IS INVOLVED IN
 BINDING TO CELL MEMBRANES.
 CC -1- SUBUNIT: CONTAINS 3 KINDS OF CHAINS. AN ALPHA AND A GAMMA CHAIN
 (FROM THE SAME PRECURSOR MOLECULE), LINKED BY AN INTERCHAIN
 DISULFIDE BOND. ASSOCIATE NONCOVALENTLY WITH AN AGGREGATE OF 4 TO
 6 BETA CHAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X00171; CAA24996.1; -
 DR EMBL: K01170; AAA27573.1; -
 DR EMBL: D30053; BAA06291.1; -
 DR EMBL: X58786; CAA41593.1; -
 DR EMBL: X76390; CAA53973.1; -
 DR EMBL: X76391; CAA53976.1; -

RESULT 10
 TENA_CHK STANDARD: PRT: 1808 AA.
 AC P10039; P13132; 073584; 073585;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Tenascin precursor (TN) (Hexabrachion) (Cytotactin) (Neuroectin)
 DE (GEM) (J1) (Mitochondrial antigen) (Glioma-associated-extracellular
 DE matrix antigen) (Gp 150-225).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=90030407; PubMed=2478295;
 RA Spring J., Beck K., Chiquet-Ehrismann R.;
 RT "Two contrary functions of tenascin: dissection of the active sites
 RT by recombinant tenascin fragments.";
 RL Cell 59:325-334(1989).
 RN [2]
 RP SEQUENCE OF 27-722 FROM N.A., AND SEQUENCE OF 79-96.
 RC TISSUE=Fibroblast;
 RX MEDLINE=89030589; PubMed=2460335;
 RA Pearson C.A., Pearson D., Shibahara S., Hofsteenge J.,
 RA Chiquet-Ehrismann R.;
 RT "Tenascin: cDNA cloning and induction by TGF-beta.";
 RL EMBO J. 7:2977-2982(1988).
 RN [3]
 RP SEQUENCE OF 464-1018 AND 1412-1661 FROM N.A., AND SEQUENCE OF 852-868.
 RC TISSUE=Embryo;
 RX MEDLINE=88176910; PubMed=2451243;
 RA Jones F.S., Burgoon M.P., Hoffman S., Crossin K.L., Cunningham B.A.,
 RA Edelman G.M.;
 RT "A cDNA clone for cytotactin contains sequences similar to epidermal
 RT growth factor-like repeats and segments of fibronectin and
 RT fibronogen.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:2186-2190(1988).
 CC -1- FUNCTION: SAM (SUBSTRATE-ADHESION MOLECULE) THAT APPEARS TO
 CC INHIBIT CELL MIGRATION. MAY PLAY A ROLE IN SUPPORTING THE GROWTH
 CC OF EPITHELIAL TUMORS. IS A LIGAND FOR INTEGRINS ALPHA-8/BETA-1,
 CC ALPHA-9/BETA-1, ALPHA-V/BETA-3 AND ALPHA-V/BETA-6.
 CC -1- SUBUNIT: HEXAMERIC. A HOMOTRIMER MAY BE FORMED IN THE TRIPLE
 CC COILED-COIL REGION AND MAY BE STABILIZED BY DISULFIDE RINGS AT
 CC BOTH ENDS. TWO OF SUCH HALF-HEXABRACHIONS MAY BE DISULFIDE LINKED
 CC WITHIN THE CENTRAL GLOBULE.
 CC -1- SUBCELLULAR LOCATION: Extracellular matrix.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS, 230 KDA (SHOWN HERE), 200 KDA
 CC AND 190 KDA, ARE PRODUCED BY ALTERNATIVE SPLICING. THEY ARE
 CC PRODUCED IN A TISSUE- AND TIME-SPECIFIC MANNER DURING DEVELOPMENT.
 CC -1- INDUCTION: BY TGF-BETA.
 CC -1- SIMILARITY: CONTAINS 13.5 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 11 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS A DOMAIN RELATED TO THE C-TERMINAL PART OF
 CC THE BETA AND GAMMA CHAINS OF FIBRINOGEN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M23121; AAA49086.1; -;
 DR EMBL: X08031; CAB40811.1; -;
 DR EMBL: X08030; CAB30824.1; ALT_TERM.
 DR EMBL: J03641; AAA48748.1; ALT_SEQ.
 DR EMBL: M20816; AAA48749.1; ALT_SEQ.
 DR PIR: A30903; A30903.

DR PIR: A31930; A31930.
 DR PIR: A33379; A33379.
 DR PIR: B33379; B33379.
 DR PIR: C33379; C33379.
 DR PIR: S01292; S01292.
 DR HSSP: P24821; IYEN.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR003964; FN.III.
 DR InterPro: IPR002181; Fibronogen_C.
 DR Pfam: PF00006; EGF_13.
 DR Pfam: PF00147; fibronogen_C; 1.
 DR Pfam: PF00041; fn3; 11.
 DR SMART: SM00181; EGF_10.
 DR SMART: SM00001; EGF-like; 1.
 DR SMART: SM00186; FBG; 1.
 DR SMART: SM00060; FN3; 10.
 DR PROSITE: PS00022; EGF_1; 14.
 DR PROSITE: PS01186; EGF_2; 14.
 KW Glycoprotein; Cell adhesion; Repeat; EGF-like domain; Coiled coil;
 KW Extracellular matrix; Signal; Alternative splicing.
 FT SIGNAL 1 22
 FT PROPEP 23 33
 FT CHAIN 34 1808
 FT DOMAIN 118 142
 FT DOMAIN 176 188
 FT DOMAIN 188 219
 FT DOMAIN 219 250
 FT DOMAIN 250 281
 FT DOMAIN 281 312
 FT DOMAIN 312 343
 FT DOMAIN 343 374
 FT DOMAIN 374 405
 FT DOMAIN 405 436
 FT DOMAIN 436 467
 FT DOMAIN 467 498
 FT DOMAIN 498 529
 FT DOMAIN 529 560
 FT DOMAIN 560 591
 FT DOMAIN 591 680
 FT DOMAIN 681 771
 FT DOMAIN 772 862
 FT DOMAIN 863 954
 FT DOMAIN 955 1042
 FT DOMAIN 1043 1133
 FT DOMAIN 1134 1224
 FT DOMAIN 1225 1315
 FT DOMAIN 1316 1404
 FT DOMAIN 1405 1492
 FT DOMAIN 1493 1580
 FT DOMAIN 1589 1808
 FT DISULFID 64 202
 FT DISULFID 192 207
 FT DISULFID 196 207
 FT DISULFID 209 218
 FT DISULFID 223 233
 FT DISULFID 227 238
 FT DISULFID 240 249
 FT DISULFID 244 264
 FT DISULFID 254 269
 FT DISULFID 258 280
 FT DISULFID 271 280
 FT DISULFID 285 295
 FT DISULFID 289 300
 FT DISULFID 302 311
 FT DISULFID 316 326
 FT DISULFID 320 331
 FT DISULFID 320 342
 FT DISULFID 333 347
 FT DISULFID 347 357
 FT DISULFID 351 362
 FT DISULFID 364 373
 FT DISULFID 378 388
 FT DISULFID 382 393
 FT DISULFID 395 404
 FT DISULFID 409 419

```

FT DISULFID 413 424 BY SIMILARITY.
FT DISULFID 426 435 BY SIMILARITY.
FT DISULFID 440 450 BY SIMILARITY.
FT DISULFID 444 455 BY SIMILARITY.
FT DISULFID 457 466 BY SIMILARITY.
FT DISULFID 471 481 BY SIMILARITY.
FT DISULFID 475 486 BY SIMILARITY.
FT DISULFID 488 497 BY SIMILARITY.
FT DISULFID 502 512 BY SIMILARITY.
FT DISULFID 506 517 BY SIMILARITY.
FT DISULFID 519 528 BY SIMILARITY.
FT DISULFID 533 543 BY SIMILARITY.
FT DISULFID 537 548 BY SIMILARITY.
FT DISULFID 550 559 BY SIMILARITY.
FT DISULFID 564 574 BY SIMILARITY.
FT DISULFID 568 579 BY SIMILARITY.
FT DISULFID 581 590 BY SIMILARITY.
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 328 328 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 751 751 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 759 759 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1050 1050 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1090 1090 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1101 1101 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1112 1112 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1153 1153 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1183 1183 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1416 1416 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1736 1736 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1769 1769 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1043 1234 MISSING (IN 200 KDA ISOFORM).
FT VARSPPLIC 1043 1315 MISSING (IN 190 KDA ISOFORM).
FT VARSPPLIC 182 182 W -> R (IN REF. 2).
FT CONFLICT 563 571 SCPNDCNNV -> PAVPTWTW (IN REF. 3).
FT CONFLICT 598 598 E -> G (IN REF. 3).
FT CONFLICT 838 838 T -> TEY (IN REF. 3).
FT CONFLICT 886 886 N -> F (IN REF. 3).
SO SEQUENCE 1808 AA; 198858 MW; B924A06CF9E9D6DE CRC64;

```

```

Query Match 40.2%; Score 43; DB 1; Length 1808;
Best Local Similarity 58.3%; Pred. No. 93;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

```
Qy 2 ETRQVEVPGSOH 13
```

```
Db 1087 EFWNITVPGSOH 1098
```

```

RESULT 11
ST14_HUMAN STANDARD; PRT; 855 AA.
AC Q9Y5Y6; Q9H3S0; Q9HCA3; Q9BS01; Q9HB36;
DT 16-OCT-2001 (Rel. 40; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DE Suppressor of tumorigenicity 14 (bc 3.4.21.-) (Matrilipase) (Membrane-
DE type serine protease 1) (MT-SPI) (Prostanin) (Serine protease TADG-15)
DE (Tumor associated differentially-expressed gene-15 protein).
GN ST14 OR PRSS14 OR SMC19 OR TADG15.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RA MEDLINE=99303581; PubMed=10373424;
RA Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;
RT "Molecular cloning of cDNA for matrilipase, a matrix-degrading serine

```

```

RT protease with trypsin-like activity."
RL J. Biol. Chem. 274:18231-18236(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99432178; PubMed=10500122;
RA Takeuchi T., Shuman M.A., Craik C.S.;
RT "Reverse biochemistry: use of macromolecular protease inhibitors to
RT dissect complex biological processes and identify a membrane-type
RT serine protease in epithelial cancer and normal tissue."
RL Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Yamaguchi N., Mitsui S.;
RT "Molecular cloning of a novel transmembrane serine protease expressed
RT in human prostate."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Tanimoto H., Underwood L.J., Wang Y., Shigemasa K., Parmley T.H.,
RT O'Brien T.J.;
RT Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 327-855 FROM N.A.
RC TISSUE=Muscle;
RA Strassberg R.;
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 340-664 FROM N.A.
RA Cao J., Fan W., Zheng S.;
RT "Genomic analysis of a novel human serine protease SMC1."
RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP CHARACTERIZATION.
RC TISSUE=Milk;
RA MEDLINE=99303582; PubMed=10373425;
RX Lin C.Y., Anders J., Johnson M., Dickson R.B.;
RT "Purification and characterization of a complex containing matrilipase
RT and a kunitz-type serine protease inhibitor from human milk."
RL J. Biol. Chem. 274:18237-18242(1999).
CC -!- FUNCTION: DEGRADERS EXTRACELLULAR MATRIX. PROPOSED TO PLAY A ROLE
CC IN BREAST CANCER INVASION AND METASTASIS. EXHIBITS TRYPSIN-LIKE
CC ACTIVITY AS DEFINED BY CLEAVAGE OF SYNTHETIC SUBSTRATES WITH ARG
CC OR LYS AS THE P1 SITE.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF118224; AADA2765.2; -
DR EMBL: AF133086; AAF00109.1; -
DR EMBL: AB030036; BAB20376.1; -
DR EMBL: AF057145; AAG15395.1; -
DR EMBL: BC005826; AAH05826.1; -
DR EMBL: AF283256; AAG13949.1; -
DR HSSP: P00763; IDPO.
DR MEROPS: S01.302; -.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR002172; LDL_recept_A.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00057; LDL_recept_a; 4.
DR Pfam: PF00089; trypsin; 1.

```

DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00192; LDLA; 3.
 DR SMART; SM00020; TRYP_SPE; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01209; LDLRA_1; 2.
 DR PROSITE; PS50068; LDLRA_2; 4.
 DR PROSITE; PS50240; TRYP SIN_DOM; 1.
 DR PROSITE; PS00134; TRYP SIN_HIS; 1.
 DR PROSITE; PS00135; TRYP SIN_SER; 1.
 KW Signal anchor; Glycoprotein; Hydrolase; Serine protease;
 KW Transmembrane; Repeat.
 FT DOMAIN 1 55 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 56 76 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT DOMAIN 77 85 (POTENTIAL).
 FT DOMAIN 214 334 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 340 447 CUB 1.
 FT DOMAIN 452 487 LDL-RECEPTOR CLASS A 1.
 FT DOMAIN 487 524 LDL-RECEPTOR CLASS A 2.
 FT DOMAIN 524 560 LDL-RECEPTOR CLASS A 3.
 FT DOMAIN 566 603 LDL-RECEPTOR CLASS A 4.
 FT DOMAIN 615 854 SERINE PROTEASE.
 FT ACT_SITE 656 656 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 711 711 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 805 805 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 302 302 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 772 772 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 327 329 FEA -> GNR (IN REF. 5).
 FT CONFLICT 381 381 R -> S (IN REF. 4).
 FT CONFLICT 674 674 A -> V (IN REF. 3).
 SQ SEQUENCE 855 AA; 94769 MW; 26143132C01F99C9 CRC64;

Query Match 39.3%; Score 42; DB 1; Length 855;
 Best Local Similarity 50.0%; Pred. No. 61;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 3 TFOVEVPGSQHI 14
 I: |||: ||:
 DB 367 TWNIEVPNNQHV 378

RESULT 12
 AROC_YEAST STANDARD; PRT; 376 AA.
 ID AROC_YEAST
 AC P28777:
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Chorismate synthase (EC 4.6.1.4) (5-enolpyruvylshikimate-3-phosphate
 DE phosphoyase).
 GN ARO2 OR YG1148V.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C;
 RX MEDLINE=92114793; PubMed=1837329;
 RA Jones D.G.U., Reusser U., Braus G.H.;
 RT "Molecular cloning, characterization and analysis of the regulation
 RT of the ARO2 gene, encoding chorismate synthase, of Saccharomyces
 RT cerevisiae.",
 RL Mol. Microbiol. 5:2143-2152(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C / FY1769;
 RX MEDLINE=97197983; PubMed=9046099;

RA Voet M., Defoor E., Verhasselt P., Riles L., Robben J., Volckaert G.;
 RT "The sequence of a nearly unclonable 22.8 kb segment on the left arm
 RT chromosome VII from Saccharomyces cerevisiae reveals ARO2, RPL9A,
 RT TRP1, MRF1 genes and six new open reading frames";
 RL Yeast 13:177-182(1997).
 CC -1- CATALYTIC ACTIVITY: 5-O-(1-carboxyvinyl)-3-phosphoshikimate =
 CC chorismate + phosphate.
 CC -1- COFACTOR: REDUCED FLAVIN.
 CC -1- PATHWAY: SEVENTH STEP IN THE BIOSYNTHESIS OF CHORISMATE WITHIN
 CC THE BIOSYNTHESIS OF AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
 CC -1- INDUCTION: BY AMINO ACID STARVATION.
 CC -1- SIMILARITY: BELONGS TO THE CHORISMATE SYNTHASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; X60190; CAA42745.1; -;
 DR EMBL; X99960; CAA68214.1; -;
 DR EMBL; Z72670; CAA96860.1; -;
 DR PIR; S17246; S17246.
 DR SGD; S0003116; ARO2.
 DR InterPro; IPR000453; Chorismate_synth.
 DR Pfam; PF01264; Chorismate_synth; 1.
 DR ProDom; PD002941; Chorismate_synth; 1.
 DR PROSITE; PS00787; CHORISMATE SYNTHASE 1; 1.
 DR PROSITE; PS00788; CHORISMATE SYNTHASE 2; 1.
 DR PROSITE; PS00789; CHORISMATE SYNTHASE 3; 1.
 DR PROSITE; PS00789; CHORISMATE SYNTHASE 3; 1.
 KW Lyase; Aromatic amino acid biosynthesis.
 SQ SEQUENCE 376 AA; 40838 MW; AF3AF65605B91E8E CRC64;

Query Match 38.8%; Score 41.5; DB 1; Length 376;
 Best Local Similarity 62.5%; Pred. No. 31;
 Matches 10; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

OY 1 GETPQ-VEVPGSQHI 15
 I: || ||||: ||:
 DB 258 GSGFGVSPGSKHND 273

RESULT 13
 GABT_HUMAN STANDARD; PRT; 500 AA.
 ID GABT_HUMAN
 AC P80404;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE 4-aminobutyrate aminotransferase, mitochondrial precursor
 DE (EC 2.6.1.19) (Gamma-amino-N-butyrate transaminase) (GABA
 DE transaminase) (GABA aminotransferase) (GABA-AT).
 GN ABAT OR GABAT.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=95237607; PubMed=7721088;
 RA Osei Y.D., Churchich J.E.;
 RT "Screening and sequence determination of a cDNA encoding the human
 RT brain 4-aminobutyrate aminotransferase";
 RL Gene 155:185-187(1995).
 RN [2]
 RP SEQUENCE OF 368-465 FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Liver;
 RX MEDLINE=95154329; PubMed=7851425;

```

RA de Bise D., Barra D., Simmaco M., John R.A., Bossa F.;
RT "Primary structure and tissue distribution of human 4-aminobutyrate
  aminotransferase."
RL Eur. J. Biochem. 227:476-480(1995).
RN [3]
RP VARIANT LYS-220.
RX MEDLINE=99336116; PubMed=10407778;
RA Medina-Kaue L.R., Tobin A.J., De Weirleir L., Jaeken J., Jakobs C.,
  Nyhan W.L., Gibson K.M.;
RT "4-aminobutyrate aminotransferase (GABA-transaminase) deficiency."
RL J. Inher. Metab. Dis. 22:414-427(1999).
CC -1- CATALYTIC ACTIVITY: 4-aminobutanoate + 2-oxoglutarate = succinate
  semialdehyde + L-glutamate.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- SUBUNIT: MONOMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -1- TISSUE SPECIFICITY: LIVER > PANCREAS > BRAIN > KIDNEY > HEART >
  PLACENTA.
CC -1- DISEASE: DEFECTS IN ABAT ARE A CAUSE OF GABA-AT DEFICIENCY WHOSE
  PHENOTYPE INCLUDES PSYCHOMOTOR RETARDATION, HYPOTONIA,
  HYPERREFLEXIA, LETHARGY, REFRACTORY SEIZURES, AND EEG
  ANOMALITIES.
CC -1- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT
  AMINOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L32961; AAA74449.1; -.
DR HSSP: P80147; 1GTX.
DR MIM: 137150; -.
DR InterPro: IPR000954; AminoTran_3.
DR Pfam: PF00202; aminoTran_3; 1.
DR PROSITE: PS00600; AA-TRANSFER_CLASS_3; 1.
KW Transferrase; Aminotransferase; Pyridoxal phosphate;
  Neurotransmitter degradation; Mitochondrion; Transit peptide;
  Disease mutation.
KW Disease mutation.
FT TRANSIT 1 28 MITOCHONDRION.
FT CHAIN 29 500 4-AMINO-BUTYRATE AMINOTRANSFERASE.
FT BINDING 357 357 PYRIDOXAL PHOSPHATE.
FT VARIANT 220 220 R->K (IN GABA-AT DEFICIENCY; 25%
  REDUCTION IN ACTIVITY).
  /FTID=VAR_008883.
FT FT 109 109 D->H (IN REF. 2).
FT FT 113 113 V->L (IN REF. 2).
FT FT 132 132 E->G (IN REF. 2).
FT FT 191 191 K->Q (IN REF. 2).
FT FT 204 204 W->G (IN REF. 2).
FT FT 216 216 S->A (IN REF. 2).
FT FT 268 268 G->R (IN REF. 2).
FT FT 320 320 C->G (IN REF. 2).
FT FT 366 366 L->H (IN REF. 2).
SQ SEQUENCE 500 AA; 56557 MW; 41199085693F80AD CRC64;

Query Match 38.8%; Score 41.5; DB 1; Length 500;
Best Local Similarity 55.6%; Pred. No. 41;
Matches 10; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

QY 3 TFOVEVPGSGHIDSGKKA 20
  I:::IIII:IIII
DB 17 TYRLLVPGSRHI-SQAAA 33

RESULT 14
RFH_ECOLI
ID AC P28369; P77246; STANDARD; PRT; 141 AA.
DR PROSITE: PS00745; RF-1; 1.

```


KW Protein biosynthesis; Complete proteome.
 FT DOMAIN 104 107 POLY-GLN.
 FT CONFLICT 1 6 MGIRK -> MEETGRKSDTLRSALVSLDGDNMAISES
 FT (IN REF. 2 AND 4).
 SQ SEQUENCE 141 AA; 16177 MW; 9D2BD9AF7A04831 CRC64;

Query Match 38.3%; Score 41; DB 1; Length 141;
 Best Local Similarity 35.0%; Pred. No. 13;
 Matches 7; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

OY 2 ETPQVEVPGSOHIDSQKAI 21
 11 : 11 11 :
 DB 49 ETLRSSGPGQHVHKTDSAV 68

RESULT 15
 YP68_CAEEL STANDARD; PRT; 313 AA.
 ID YP68_CAEEL
 AC 009217;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 37.0 kDa protein B0495.8 in chromosome II.
 GN B0495.8.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Kirsten J.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 -I- SIMILARITY: SOME. TO YEAST YDL087C AND S.POMBE SPCC16A11.13.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; U21317; AAA62527.1; -
 DR Wormpep; B0495.8; CE01766.
 KW Hypothetical protein.
 SQ SEQUENCE 313 AA; 36977 MW; 000D2327621BFED0 CRC64;

Query Match 38.3%; Score 41; DB 1; Length 313;
 Best Local Similarity 46.2%; Pred. No. 30;
 Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 7 ETPGSOHIDSOKK 19
 :: ||||:|:::
 DB 11 QLMGSOHVDNKKR 23

Search completed: July 3, 2002, 09:04:43
 Job time: 583 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 09:06:05 ; Search time 74.17 Seconds
(without alignments)
48.981 Million cell updates/sec

Title: US-09-786-648-5
Perfect score: 107
Sequence: 1 GETFQVEVPGSHIDISQKKAI 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	94.4	124	2	057193
2	101	94.4	124	2	09RP15
3	101	94.4	124	2	056635
4	101	94.4	124	2	093V32
5	101	94.4	124	9	094M01
6	95	88.8	103	2	09R646
7	54	50.5	108	4	099987
8	52	48.6	109	10	09LJBO
9	49	45.8	255	10	09C815
10	49	45.8	257	10	09C875
11	49	45.8	428	10	09LGM2
12	47	43.9	392	10	040742
13	47	43.9	395	5	09NRD5
14	46	43.0	91	10	09SYF4
15	46	43.0	574	10	09SXC9
16	45.5	42.5	1166	3	09PA60

17	45	42.1	439	10	09LY67	091Y67 arabidopsis
18	45	42.1	2732	12	09J3F2	09J3F2 murine hepa
19	45	42.1	2733	12	09PR42	09PR42 murine hepa
20	45	42.1	2733	12	039226	039226 murine hepa
21	45	42.1	2733	12	09J3E8	09J3E8 murine hepa
22	44.5	41.6	565	10	022511	022511 vitis vinif
23	44	41.1	293	10	049876	049876 lupinus alb
24	44	41.1	330	11	09QZL7	09QZL7 mus musculu
25	44	41.1	374	5	09U3M9	09U3M9 caenorhabdi
26	44	41.1	467	6	09N136	09N136 ovis aries
27	44	41.1	467	11	091VD0	091VD0 mus musculu
28	44	41.1	918	3	09P606	09P606 neurospora
29	43.5	40.7	641	10	09SKB2	09SKB2 arabidopsis
30	43.5	40.7	641	10	093Z40	093Z40 arabidopsis
31	43	40.2	356	5	09NF90	09NF90 leishmania
32	43	40.2	368	10	09SN44	09SN44 arabidopsis
33	43	40.2	843	12	09QMN7	09QMN7 hepatitis b
34	43	40.2	843	12	091817	091817 hepatitis b
35	43	40.2	849	16	0981D6	0981D6 rhizobium l
36	43	40.2	1714	13	090995	090995 gallus gall
37	43	40.2	1810	13	090824	090824 gallus gall
38	42	39.3	204	16	09H7A0	09H7A0 pseudomonas
39	42	39.3	330	16	099VM1	099VM1 staphylococ
40	42	39.3	427	13	042099	042099 cyprinus ca
41	42	39.3	672	16	09A818	09A818 caulobacter
42	42	39.3	697	5	0965W6	0965W6 caenorhabdi
43	42	39.3	802	10	0947W6	0947W6 oryza sativ
44	42	39.3	945	10	094D27	094D27 oryza sativ
45	42	39.3	946	10	094E00	094E00 oryza sativ

ALIGNMENTS

RESULT 1
057193 PRELIMINARY; PRT; 124 AA.
AC 057193;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CHOLERA TOXIN B PROTEIN (CTB) PRECURSOR (CTB).
GN CTXB.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLASSICAL STRAIN 569B;
RX MEDLINE=9135224; PubMed=1883840;
RA Dams E., De Wolf M., Dierick W.;
RT "Nucleotide sequence analysis of the CT operon of the Vibrio cholerae
RT classical strain 569B.";
RL Biochim. Biophys. Acta 1090:139-141(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CLASSICAL BIOTYPE 569B;
RA Shi C., Cao C., Zhang J., Ma O.;
RN Chin. Biochem. J. 9:395-399(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CLASSICAL BIOTYPE 569B;
RA Xu L.;
RT Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: X58785; CAA41591.1; -;
DR EMBL: U25679; AAC34728.1; -;
DR EMBL: A00931; CAA00098.1; -;
DR HSSP: P01536; ZCHB.
DR InterPro: IPR001835; Enterotoxin_B.
DR Pfam: PF01376; Enterotoxin_B.1.
DR PRINTS: PR00772; ENTEROTOXINB.
DR ProDom: PD012805; Enterotoxin_B.1.
KW Signal.

RESULT	5		
094M01			
ID	094M01	PRELIMINARY;	PRT; 124 AA.
AC	094M01;		
DT	01-DEC-2001	(TREMBLrel. 19, Created)	
DT	01-DEC-2001	(TREMBLrel. 19, Last sequence update)	
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)	
DE	CTXB.		
GN	CTXB.		
OS	Vibrio phage CTX.		
OC	viruses; ssDNA viruses; Inoviridae; Inovirus.		
OX	NCBI_TaxID=141904;		
NN	[1]		

2000
 2001
 2002
 2003
 2004
 2005
 2006
 2007
 2008
 2009
 2010
 2011
 2012
 2013
 2014
 2015
 2016
 2017
 2018
 2019
 2020
 2021
 2022
 2023
 2024
 2025
 2026
 2027
 2028
 2029
 2030
 2031
 2032
 2033
 2034
 2035
 2036
 2037
 2038
 2039
 2040
 2041
 2042
 2043
 2044
 2045
 2046
 2047
 2048
 2049
 2050
 2051
 2052
 2053
 2054
 2055
 2056
 2057
 2058
 2059
 2060
 2061
 2062
 2063
 2064
 2065
 2066
 2067
 2068
 2069
 2070
 2071
 2072
 2073
 2074
 2075
 2076
 2077
 2078
 2079
 2080
 2081
 2082
 2083
 2084
 2085
 2086
 2087
 2088
 2089
 2090
 2091
 2092
 2093
 2094
 2095
 2096
 2097
 2098
 2099
 2100
 2101
 2102
 2103
 2104
 2105
 2106
 2107
 2108
 2109
 2110
 2111
 2112
 2113
 2114
 2115
 2116
 2117
 2118
 2119
 2120
 2121
 2122
 2123
 2124
 2125
 2126
 2127
 2128
 2129
 2130
 2131
 2132
 2133
 2134
 2135
 2136
 2137
 2138
 2139
 2140
 2141
 2142
 2143
 2144
 2145
 2146
 2147
 2148
 2149
 2150
 2151
 2152
 2153
 2154
 2155
 2156
 2157
 2158
 2159
 2160
 2161
 2162
 2163
 2164
 2165
 2166
 2167
 2168
 2169
 2170
 2171
 2172
 2173
 2174
 2175
 2176
 2177
 2178
 2179
 2180
 2181
 2182
 2183
 2184
 2185
 2186
 2187
 2188
 2189
 2190
 2191
 2192
 2193
 2194
 2195
 2196
 2197
 2198
 2199
 2200
 2201
 2202
 2203
 2204
 2205
 2206
 2207
 2208
 2209
 2210
 2211
 2212
 2213
 2214
 2215
 2216
 2217
 2218
 2219
 2220
 2221
 2222
 2223
 2224
 2225
 2226
 2227
 2228
 2229
 2230
 2231
 2232
 2233
 2234
 2235
 2236
 2237
 2238
 2239
 2240
 2241
 2242
 2243
 2244
 2245
 2246
 2247
 2248
 2249
 2250
 2251
 2252
 2253
 2254
 2255
 2256
 2257
 2258
 2259
 2260
 2261
 2262
 2263
 2264
 2265
 2266
 2267
 2268
 2269
 2270
 2271
 2272
 2273
 2274
 2275
 2276
 2277
 2278
 2279
 2280
 2281
 2282
 2283
 2284
 2285
 2286
 2287
 2288
 2289
 2290
 2291
 2292
 2293
 2294
 2295
 2296
 2297
 2298
 2299
 2300
 2301
 2302
 2303
 2304
 2305
 2306
 2307
 2308
 2309
 2310
 2311
 2312
 2313
 2314
 2315
 2316
 2317
 2318
 2319
 2320
 2321
 2322
 2323
 2324
 2325
 2326
 2327
 2328
 2329
 2330
 2331
 2332
 2333
 2334
 2335
 2336
 2337
 2338
 2339
 2340
 2341
 2342
 2343
 2344
 2345
 2346
 2347
 2348
 2349
 2350
 2351
 2352
 2353
 2354
 2355
 2356
 2357
 2358
 2359
 2360
 2361
 2362
 2363
 2364
 2365
 2366
 2367
 2368
 2369
 2370
 2371
 2372
 2373
 2374
 2375
 2376
 2377
 2378
 2379
 2380
 2381
 2382
 2383
 2384
 2385
 2386
 2387
 2388
 2389
 2390
 2391
 2392
 2393
 2394
 2395
 2396
 2397
 2398
 2399
 2400
 2401
 2402
 2403
 2404
 2405
 2406
 2407
 2408
 2409
 2410
 2411
 2412
 2413
 2414
 2415
 2416
 2417
 2418
 2419
 2420
 2421
 2422
 2423
 2424
 2425
 2426
 2427
 2428
 2429
 2430
 2431
 2432
 2433
 2434
 2435
 2436
 2437
 2438
 2439
 2440
 2441
 2442
 2443
 2444
 2445
 2446
 2447
 2448
 2449
 2450
 2451
 2452
 2453
 2454

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
 ON NCBI_TaxID=3702;
 RX SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Cressy T.H., Dewar K.,
 RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Feng B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizer L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremetska I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Mafti R., Marziani A.,
 RA Miltischer J., Miranda M., Nguyen M., Niernan W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana."
 RL Nature 408:816-820(2000).
 DR EMBL: AC051630; AAG51209.1; -
 DR InterPro: IPR000352; Pep_rel_factor_1.
 DR Pfam: PF00472; RF-1; 1
 SO SEQUENCE 255 AA; 27445 MW; B5AF38991FF0D0F8 CRC64;

Query Match 45.8%; Score 49; DB 10; Length 255;
 Best local Similarity 45.0%; Pred. No. 3.9;
 Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 2 ETRVPSGQHIDSOKKAI 21
 DB 101 ETRVPSGQHIDKRSAY 120

RESULT 10
 OY 09C875 PRELIMINARY; PRT; 257 AA.
 AC 09C875;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE HYPOTHETICAL 27.6 KDA PROTEIN.
 GN T1609.18.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
 ON NCBI_TaxID=3702;
 RX SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Cressy T.H., Dewar K.,
 RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Feng B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizer L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremetska I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Mafti R., Marziani A.,
 RA Miltischer J., Miranda M., Nguyen M., Niernan W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,

RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana."
 RL Nature 408:816-820(2000).
 DR EMBL: AC027035; AAG51290.1; -
 DR InterPro: IPR000352; Pep_rel_factor_1.
 DR Pfam: PF00472; RF-1; 1
 KW Hypothetical protein.
 SO SEQUENCE 257 AA; 27645 MW; 9F85B09C029E97C9 CRC64;

Query Match 45.8%; Score 49; DB 10; Length 257;
 Best local Similarity 45.0%; Pred. No. 3.9;
 Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 2 ETRVPSGQHIDSOKKAI 21
 DB 101 ETRVPSGQHIDKRSAY 120

RESULT 11
 OY 09LGM2 PRELIMINARY; PRT; 428 AA.
 AC 09LGM2;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE ESTS A0056822(S20908).
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 ON NCBI_TaxID=4530;
 RX SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 clone: P0041E11."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 clone: P0433F09."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AP002521; BAA96762.1; -
 DR EMBL: AP002539; BAA08201.1; -
 DR HSSP: 006319; 1BUC.
 DR InterPro: IPR001552; Acyl-CoA-dh.
 DR Pfam: PF00441; Acyl-CoA-dh.1.
 DR Pfam: PF02770; Acyl-CoA-dh.M.1.
 DR Pfam: PF02771; Acyl-CoA-dh.N.1.
 DR PROSITE: PS00073; ACYL_COA_DH_2; UNKNOWN_1.
 SO SEQUENCE 428 AA; 46132 MW; 8D34E3698A8BE6367 CRC64;

Query Match 45.8%; Score 49; DB 10; Length 428,
 Best local Similarity 52.6%; Pred. No. 7;
 Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 3 ETRVPSGQHIDSOKKAI 21
 DB 183 TATKVPGGWHIDGQKRWI 201

RESULT 12
 ID 040742 PRELIMINARY; PRT; 392 AA.
 AC 040742;
 DT 040742;

DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE OSRAD23.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Euphorbiaceae; Oryzaeae; Oryza.
 NCBI_TaxID=4530;
 RN NCBI_TaxID=4530;
 RP SEQUENCE FROM N.A.
 RC STRAIN=NIPPONBARE;
 RX MEDLINE=97369378; PubMed=9225866;
 RA Schultz T.F., Quattrano R.S.;
 RT "Characterization and expression of a rice RAD23 gene."
 RL Plant Mol. Biol. 34:557-562(1997).
 DR EMBL: U63530; AAB65841.1; -;
 DR HSSP: P54725; IDVO.
 DR InterPro: IPR000449; UBA.
 DR InterPro: IPR000626; Ubiquitin.
 DR Pfam: PF00627; UBA; 2.
 DR Pfam: PF00240; Ubiquitin; 1.
 DR SMART: SM00165; UBA; 2.
 DR SMART: SM00213; UBO; 1.
 DR PROSITE: PS50053; UBQUITIN_2; 1.
 SQ SEQUENCE 392 AA; 41754 MW; BDE08574CC7C6B CRC64;

 Query Match 43.9%; Score 47; DB 10; Length 392;
 Best Local Similarity 42.9%; Pred. No. 14;
 Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

 OY 1 GETFVEVPGSQHDSOKRAI 21
 1 |||:| | : | : |
 DB 10 GSTFQLEVDASQKVDVKRII 30

 RESULT 13
 Q9NKK5 PRELIMINARY; PRT; 395 AA.
 AC Q9NKK5;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE HYPOTHETICAL 43.6 KDA PROTEIN.
 GN BG:DS01514.3.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydriidae; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RN NCBI_TaxID=7227;
 RP SEQUENCE FROM N.A.
 RC STRAIN=y, CN BW SP;
 RX MEDLINE=99403001; PubMed=10471707;
 RA Ashburner M., Mitra S., Roote J., Lewis S.E., Blazej R., Davis T.,
 RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
 RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
 RA Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitehead K.,
 RA Celisner S., Rubin G.M.;
 RT "An exploration of the sequence of a 2.9-Mb region of the genome of
 RT Drosophila melanogaster: the Adh region."
 RL Genetics 153:179-219(1999).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=y, CN BW SP;
 RA Celisner S.E., Agdayani A., Arcaina T.T., Baxter E., Blazej R.G.,
 RA Butenhardt C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
 RA Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
 RA Houston K.A., Hummel S.R., Kaira K., Kearney L., Kim E., Lee B.,
 RA Lewis S., Li P., Lomlati S.A., Mazda P., Moshrefi A.R., Moshrefi M.,
 RA Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Segueira A.,
 RA Sethi H., Snir E., Svirkas R.R., Wan K.H., Weinburg T., Zhang R.,

RA Zieran L.L., Rubin G.M.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AE003408; AAF44847.1; -;
 DR Flybase: FBgn0028907; BG:DS01514.3.
 KW Hypothetical protein.
 SQ SEQUENCE 395 AA; 43561 MW; AEF41CC4ADD3DA73 CRC64;

 Query Match 43.9%; Score 47; DB 5; Length 395;
 Best Local Similarity 56.2%; Pred. No. 14;
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

 OY 5 OYEVPGSQHDSOKRA 20
 1 |||:| | : | : |
 DB 329 RVSPVSTHIDADANA 344

 RESULT 14
 Q9SYF4 PRELIMINARY; PRT; 91 AA.
 AC Q9SYF4;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE F1511.6 PROTEIN.
 GN F1511.6.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucotsid II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RN NCBI_TaxID=3702;
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, COLUMBIA;
 RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
 RA Li J., Kremetskaia I., Luros J., Ngan I., Gonzalez A., Altati H.,
 RA Araujo R., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,
 RA Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
 RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
 RT "Arabidopsis thaliana chromosome 1 BAC F1511 sequence."
 RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AC006577; AAD25769.1; -;
 DR HSSP: P02248; UBI.
 DR InterPro: IPR000626; Ubiquitin.
 DR Pfam: PF00240; Ubiquitin; 2.
 SQ SEQUENCE 91 AA; 10142 MW; E8766823D6450267 CRC64;

 Query Match 43.0%; Score 46; DB 10; Length 91;
 Best Local Similarity 47.6%; Pred. No. 3.8;
 Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

 OY 1 GETFVEVPGSQHDSOKRAI 21
 1 |||:| | : | : |
 DB 58 GTFNLEVGSELIQVKNMI 78

 RESULT 15
 Q9SXC9 PRELIMINARY; PRT; 574 AA.
 AC Q9SXC9;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE T17H3.2 PROTEIN (AT1G27520/T17H3_2).
 GN T17H3.2.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucotsid II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RN NCBI_TaxID=3702;
 RP SEQUENCE FROM N.A.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: July 3, 2002, 08:50:10 ; Search time 94.52 Seconds
(without alignments)
8.226 Million cell updates/sec

Title: US-09-786-648-2
Perfect score: 39
Sequence: 1 EVPGSQH 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

```
Searched:      747574 segs, 111073796 residues
Total number of hits satisfying chosen parameters:  747574
```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*

1:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA1980.DAT.*
2:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA1981.DAT.*
3:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA1982.DAT.*
4:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA1983.DAT.*
5:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA1984.DAT.*
6:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA1985.DAT.*
7:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA1986.DAT.*
8:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA1987.DAT.*
9:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA1988.DAT.*
10:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA1989.DAT.*
11:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA1990.DAT.*
12:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA1991.DAT.*
13:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA1992.DAT.*
14:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA1993.DAT.*
15:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA1994.DAT.*
16:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA1995.DAT.*
17:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA1996.DAT.*
18:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA1997.DAT.*
19:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA1998.DAT.*
20:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA1999.DAT.*
21:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA2000.DAT.*
22:	/SIDS1/gcgdata/hold-genseseq/geneseqp-emb1/AA2001.DAT.*

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	39	100.0	7	21	AA87460	Cholera toxin B/e
2	39	100.0	8	21	AA87464	Cholera toxin B/e
3	39	100.0	12	21	AA87461	Cholera toxin B/e
4	39	100.0	15	10	AA893498	CTp3 epitope of t
5	39	100.0	15	16	AA885125	Cholera toxin B a
6	39	100.0	21	21	AA87462	Cholera toxin B/e
7	39	100.0	21	21	AA87463	Cholera toxin B/e
8	39	100.0	23	16	AA876748	E. coli heat labile
9	39	100.0	26	4	AA830265	Residues 50-64 of
10	39	100.0	41	6	AA850439	Sequence of amino
11	39	100.0	46	6	AA850436	Network polymer wh

12	39	100.0	47	4	AAP30600	Sequence of amino
13	39	100.0	93	16	AAK72545	ADP-ribosylating t
14	39	100.0	93	20	AAK11816	Escherichia coli v
15	39	100.0	93	20	AAW95226	E. coli heat-labile
16	39	100.0	93	21	AAK68365	Heat labile toxin
17	39	100.0	93	22	AAK66239	E coli verotoxin-1
18	39	100.0	103	6	AAP50340	Sequence of subunit
19	39	100.0	103	17	AAW04857	Synthetic cholera
20	39	100.0	103	17	AAK94939	Heat labile enteric
21	39	100.0	103	17	AAW06060	Cholera toxin B su
22	39	100.0	103	17	AAW06607	Cholera toxin B su
23	39	100.0	103	19	AAW080808	Amino acid sequenc
24	39	100.0	103	22	AAK62365	v. cholera cholera
25	39	100.0	103	22	AAK62367	v. cholera cholera
26	39	100.0	103	22	AAK62370	v. cholera cholera
27	39	100.0	103	22	AAK62379	E. coli LTB proteol
28	39	100.0	105	22	AAK62369	v. cholera cholera
29	39	100.0	118	11	AAK04163	Cholera Toxin B-su
30	39	100.0	123	22	AAK62374	E. coli LTB proteol
31	39	100.0	123	22	AAK62377	E. coli LTB proteol
32	39	100.0	124	10	AAP93561	B subunit of the h
33	39	100.0	124	17	AAW06065	Cholera toxin B su
34	39	100.0	124	19	AAW59770	Amino acid sequenc
35	39	100.0	124	21	AAK96652	Plant-optimized E
36	39	100.0	124	21	AAK96872	Plant-optimized v
37	39	100.0	124	22	AAK65992	Cholera toxin B su
38	39	100.0	124	22	AAK62359	v. cholera strain
39	39	100.0	124	22	AAK62363	v. cholera cholera
40	39	100.0	124	22	AAK62368	v. cholera cholera
41	39	100.0	124	22	AAK62375	E. coli LTB proteol
42	39	100.0	124	22	AAK62376	E. coli LTB proteol
43	39	100.0	124	22	AAK62376	E. coli LTB proteol
44	39	100.0	124	22	AAK62376	E. coli LTB proteol
45	39	100.0	124	22	AAK62380	E. coli LTB proteol

ALIGNMENTS

PT	RESULT	1
XX	AAV87460	
XX	ID	AAV87460 standard; peptide; 7 AA.
XX	AC	
XX	AAV87460;	
DT	03-JUL-2000	(first entry)
XX		
DE	Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:2.	
XX		
KW	Cholera toxin subunit B; CTxB; heat labile enterotoxin subunit B; EtxB; beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation; adjuvant; immune disorder; diarrhoea.	
XX		
OS	Vibrio cholerae.	
OS	Escherichia coli.	
XX		
PN	WO200014114-A1.	
PD		
PD	16-MAR-2000.	
XX		
FE	07-SEP-1999; 99MO-GB02970.	
XX		
PR	07-SEP-1998; 98GB-0019484.	
XX		
RA	(UYBR-) UNIV BRISTOL.	
XX		
PI	Williams NA, Hirst TR;	
DR	WPI; 2000-256943/22.	
XX		
PT	Derivatives of Escherichia coli heat labile enterotoxins useful as immunomodulators and for treating diarrhea and which do not bind the glycolipid receptor GM-1 -	

XX Claim 1; Page 13; 62pp; English.
PS
XX The invention relates to peptide fragments of the Escherichia coli heat
CC labile enterotoxin (Etx) and its closely related homologue, cholera
CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous
CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are
CC composed of one A subunit and five identical B subunits. The A subunit
CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-
CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)
CC facilitate the entry of subunit A into the host cell via the binding and
CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible
CC for some of the effects of Etx and Ctx, it has been found that certain
CC effects of the toxins, such as immunomodulation, are not mediated
CC through GM-1 binding. The peptides of the invention are fragments of the
CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as
CC normal EtxB and CtxB subunits, except that they do not bind or cross link
CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.
CC They may also be used as an inhibitor for toxin-induced diarrhoea.
CC Therefore, the peptides may be used in the production of a composition
CC for treating, preventing and/or modulating a disease associated with an
CC immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463
CC represent preferred peptides of the invention, AAY87460 being
CC particularly preferred.
XX
SQ Sequence 7 AA;

Query Match 100.0%; Score 39; DB 21; Length 7;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSQH 7
| | | | |
Db 1 evpgsqh 7

RESULT 2
AAY87464
ID AAY87464 standard; peptide; 8 AA.
XX

AC AAY87464;

DT 03-JUL-2000 (first entry)

DE Cholera toxin B/enterotoxin B-derived immunomodulatory peptide.

KM Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;

KW beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;

KM adjuvant; immune disorder; diarrhoea.

OS Vibrio cholerae.

OS Escherichia coli.

PN WO200014114-A1.

PD 16-MAR-2000.

PF 07-SEP-1999; 99WO-GB02970.

PR 07-SEP-1998; 98GB-0019484.

PA (UYBR-) UNIV BRISTOL.

PI Williams NA, Hirst TR;

DR WPI; 2000-256943/22.

XX Derivatives of Escherichia coli heat labile enterotoxins useful as
PT immunomodulators and for treating diarrhoea and which do not bind the
PT glycolipid receptor GM-1 -
XX
PS Example 5; Page 45; 62pp; English.

XX The invention relates to peptide fragments of the Escherichia coli heat
CC labile enterotoxin (Etx) and its closely related homologue, cholera
CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous
CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are
CC composed of one A subunit and five identical B subunits. The A subunit
CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-
CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)
CC facilitate the entry of subunit A into the host cell via the binding and
CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible
CC for some of the effects of Etx and Ctx, it has been found that certain
CC effects of the toxins, such as immunomodulation, are not mediated
CC through GM-1 binding. The peptides of the invention (AAY87461-Y87463) are
CC fragments of the beta-4-alpha-2 loop of EtxB and/or CtxB, exerting the
CC same effects as normal EtxB and CtxB subunits, except that they do not
CC bind or cross link GM-1. They may be used in medicine as an
CC immunomodulator or adjuvant. They may also be used as an inhibitor for
CC toxin-induced diarrhoea. Therefore, the peptides may be used in the
CC production of a composition for treating, preventing and/or modulating a
CC disease associated with an immune disorder and/or toxin-induced
CC diarrhoea. Sequences AAY87464-Y87465 represent peptides used in an
CC exemplification of the present invention to assess whether a peptide
CC corresponding to a portion of the beta-4-alpha-2 loop of EtxB and CtxB
CC has immunomodulatory effects. Peptide AAY87464 corresponds to residues
CC 51-58 of the EtxB/CtxB beta-4-alpha-2 loop, and peptide AAY87465 is a
CC randomly selected control peptide.
XX
SQ Sequence 8 AA;

Query Match 100.0%; Score 39; DB 21; Length 8;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSQH 7
| | | | |
Db 1 evpgsqh 7

RESULT 3
AAY87461
ID AAY87461 standard; peptide; 12 AA.
XX

AC AAY87461;

DT 03-JUL-2000 (first entry)

DE Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:3.

KM Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;

KW beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;

KM adjuvant; immune disorder; diarrhoea.

OS Vibrio cholerae.

OS Escherichia coli.

PN WO200014114-A1.

PD 16-MAR-2000.

PF 07-SEP-1999; 99WO-GB02970.

PR 07-SEP-1998; 98GB-0019484.

PA (UYBR-) UNIV BRISTOL.

PI Williams NA, Hirst TR;

XX WPI; 2000-256943/22.
DR
XX Derivatives of Escherichia coli heat labile enterotoxins useful as
PT immunomodulators and for treating diarrhoea and which do not bind the
PT glycolipid receptor GM-1 -
XX
PS

XX Disclosure: Page 15; 62pp; English.
PS
XX
CC The invention relates to peptide fragments of the Escherichia coli heat
CC labile enterotoxin (Etx) and its closely related homologue, cholera
CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous
CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are
CC composed of one A subunit and five identical B subunits. The A subunit
CC is responsible for toxicity, possessing adenosine diphosphate (ADP)-
CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)
CC facilitate the entry of subunit A into the host cell via the binding and
CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible
CC for some of the effects of Etx and Ctx, it has been found that certain
CC effects of the toxins, such as immunomodulation, are not mediated
CC through GM-1 binding. The peptides of the invention are fragments of the
CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as
CC normal EtxB and CtxB subunits, except that they do not bind or cross link
CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.
CC They may also be used as an inhibitor for toxin-induced diarrhoea.
CC Therefore, the peptides may be used in the production of a composition
CC for treating, preventing and/or modulating a disease associated with an
CC immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463
CC represent preferred peptides of the invention, AAY87460 being
CC particularly preferred.
SQ Sequence 12 AA:

Query Match 100.0%; Score 39; DB 21; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.29; 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

OY 1 EYPGSGH 7
| | | | | | | |
DB 2 evpgsqh 8

RESULT 4
AAP93498
ID AAP93498 standard; protein; 15 AA.
XX
AC AAP93498;
XX
DT 03-MAY-1990 (first entry)
XX
DE CTP3 epitope of the Cholera toxin B subunit.
XX
KW CTP3 epitope of Cholera toxin B subunit; flagellin fusion protein;
KM vaccine; immunotherapy: ds;
XX
PN WO8910967-A.
XX
PD 16-NOV-1989.
XX
PF 05-MAY-1989; 89WO-US01932.
XX
PR 05-MAY-1988; 88US-0190570.
XX
PA (PRAX-) PRAXIS BIOLOGICS INC.
PA (STRD) LEYLAND STANDFORD JUNIOR UNIV.
XX
PI Marjarian WR, Stocker BAD, Newton SMC;
XX
DR WPI; 1989-356496/48.
DR N-PSDB; AAN92414.
XX
PT New recombinant flagellin gene including sequence - for heterologous
PT epitope, and expressed fusion proteins, useful in vaccines and for prodn.
PT of antibodies:
XX
PS Disclosure; fig.4B; 137pp; English.
XX This sequence corresponds to the CTP3 epitope of the Cholera toxin B

CC subunit. The DNA sequence encoding this ligates to othersynthetic
CC oligonucleotides to form a new recombinant gene. This encodes
CC a flagellin fusion protein which can be used in vaccines for immuno-
CC therapy.
XX
SQ Sequence 15 AA:

Query Match 100.0%; Score 39; DB 10; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYPGSGH 7
| | | | | | | |
DB 2 evpgsqh 8

RESULT 5
AAR85125
ID AAR85125 standard; peptide; 15 AA.
XX
AC AAR85125;
XX
DT 13-JUN-1996 (first entry)
XX
DE Cholera toxin B antigenic peptide fragment CTP3.
XX
KW Conjugate; cholera; B toxin; peptide fragment; microparticulate;
KW inert carrier; modified silica; thyroglobulin; oral vaccine;
KW immunisation; infection; insoluble; digestive tract; antigen;
KW intestines; antibodies; secretory; IGA class.
XX
OS Vibrio cholerae.
XX
PN WO9529701-A1.
XX
PD 09-NOV-1995.
XX
PF 02-MAY-1995; 95WO-EP01661.
XX
PR 03-MAY-1994; 94IL-0109519.
XX
PA (YEDA) YEDA RES & DEV CO LTD.
XX
PI Marks RS, Mirelman D, Sela M;
XX
DR WPI; 1995-403805/51.
XX
DT Vaccines for oral immunisation against infecting agents, e.g.
PT cholera - comprise a conjugate of an antigen of an infecting agent
PT covalently bound to micro:particulate inert carrier, e.g. modified
PT aldehyde silica
XX
PS Claim 7; Page 25; 40pp; English.
XX
CC A compsn. comprising a conjugate of an antigenic cholera B toxin
CC peptide fragment (e.g. AAR85126-30 and esp. AAR85125), covalently
CC bound to a microparticulate inert carrier (e.g. modified silica or
CC thyroglobulin) can be used as an oral vaccine for immunisation
CC against cholera infection. The inert carrier is insoluble in the
CC digestive tract, allowing presentation of the antigen in the
CC intestines, where it will elicit antibodies mainly of the
CC secretory IGA class.
XX
SQ Sequence 15 AA:

Query Match 100.0%; Score 39; DB 16; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYPGSGH 7
| | | | | | | |

Db 2 evpsqgh 8

RESULT 6

AA87462
ID AAY87462 standard; peptide: 21 AA.

XX AAY87462;

XX 03-JUL-2000 (first entry)

XX Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:4.

XX Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;
XX beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;
XX adjuvant; immune disorder; diarrhoea.XX Vibrio cholerae;
XX Escherichia coli.

XX WO200014114-A1.

XX 16-MAR-2000.

XX 07-SEP-1999; 99WO-G802970.

XX 07-SEP-1998; 98GB-0019484.

XX (UYBR-) UNIV BRISTOL.

XX Williams NA, Hirst TR;

XX WPI: 2000-256943/22.

XX Derivatives of Escherichia coli heat labile enterotoxins useful as
PT immunomodulators and for treating diarrhea and which do not bind the
PT glycolipid receptor GM-1 -

XX Disclosure; Page 15; 62pp; English.

XX The invention relates to peptide fragments of the Escherichia coli heat
CC labile enterotoxin (Etx) and its closely related homologue, cholera
CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous
CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are
CC composed of one A subunit and five identical B subunits. The A subunit
CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-
CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)
CC facilitate the entry of subunit A into the host cell via the binding and
CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible
CC for some of the effects of Etx and Ctx, it has been found that certain
CC effects of the toxins, such as immunomodulation, are not mediated
CC through GM-1 binding. The peptides of the invention are fragments of the
CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as
CC normal EtxB and CtxB subunits, except that they do not bind or cross link
CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.
CC They may also be used as an inhibitor for toxin-induced diarrhoea.
CC Therefore, the peptides may be used in the production of a composition
CC for treating, preventing and/or modulating a disease associated with an
CC immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463
CC represent preferred peptides of the invention, AAY87460 being
CC particularly preferred.

XX Sequence 21 AA;

Query Match 100.0%; Score 39; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPSQGH 7
Db 7 evpsqgh 13

RESULT 7

AA87463
ID AAY87463 standard; peptide: 21 AA.

XX AAY87463;

XX 03-JUL-2000 (first entry)

XX E. coli heat labile enterotoxin B-derived peptide, SPQ ID NO:5.

XX Heat labile enterotoxin subunit B; EtxB;
XX beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;
XX adjuvant; immune disorder; diarrhoea.

XX Escherichia coli.

XX WO200014114-A1.

XX 16-MAR-2000.

XX 07-SEP-1999; 99WO-G802970.

XX 07-SEP-1998; 98GB-0019484.

XX (UYBR-) UNIV BRISTOL.

XX Williams NA, Hirst TR;

XX WPI: 2000-256943/22.

XX Derivatives of Escherichia coli heat labile enterotoxins useful as
PT immunomodulators and for treating diarrhea and which do not bind the
PT glycolipid receptor GM-1 -

XX Disclosure; Page 15; 62pp; English.

XX The invention relates to peptide fragments of the Escherichia coli heat
CC labile enterotoxin (Etx) and its closely related homologue, cholera
CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous
CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are
CC composed of one A subunit and five identical B subunits. The A subunit
CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-
CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)
CC facilitate the entry of subunit A into the host cell via the binding and
CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible
CC for some of the effects of Etx and Ctx, it has been found that certain
CC effects of the toxins, such as immunomodulation, are not mediated
CC through GM-1 binding. The peptides of the invention are fragments of the
CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as
CC normal EtxB and CtxB subunits, except that they do not bind or cross link
CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.
CC They may also be used as an inhibitor for toxin-induced diarrhoea.
CC Therefore, the peptides may be used in the production of a composition
CC for treating, preventing and/or modulating a disease associated with an
CC immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463
CC represent preferred peptides of the invention, AAY87460 being
CC particularly preferred.

XX Sequence 21 AA;

Query Match 100.0%; Score 39; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPSQGH 7
Db 7 evpsqgh 13RESULT 8
AAR76748

ID AAR76748 standard; Protein; 23 AA.
 XX AAR76748;
 AC
 DT 18-MAR-1996 (first entry)
 XX
 DE Residues 50-64 of cholera toxin B subunit and FimH 224-226.
 XX
 KW FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
 KW FimA; FimF; FimG; receptor binding site; PCR; amplifly; ss.
 XX
 OS Chimeric - Vibrio cholerae.
 OS Chimeric - Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1 /note= "Represents FimH residue 224"
 FT Peptide 2..4 /note= "Linker peptide"
 FT Peptide 5..19 /note= "Cholera toxin B subunit 50-64"
 FT Peptide 20..22 /note= "Linker peptide"
 FT Misc-difference 23 /note= "Represents FimH residue 226"
 FT
 XX W09520657-A1.
 PN
 PD 03-AUG-1995.
 XX
 PF 27-JAN-1995; 95WO-DK00042.
 XX
 PR 27-JAN-1994; 94US-0187166.
 XX
 PA (GXBI-) GX BIOSYSTEMS AS.
 XX
 PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
 XX
 DR WPI; 1995-275442/36.
 DR N-PSDB; AAQ93061.
 XX
 PT Receptor specific bacterial adhesins - useful for targetting active
 PT compounds and microbial cells to locations of receptors
 PS
 PS Example 3; Page 58; 152pp; English.
 XX
 CC This sequence is encoded by a fragment of the the plasmid pLP93
 CC which was used in the production of fimH fusion genes comprising
 CC the cholera toxin B subunit inserted into the fimH gene. This insert
 CC shows the inclusion of the B subunit into the FimH protein at position
 CC 224-226. The chimeric genes were then opt. further modified by insertion
 CC of the hepatitis B virus surface antigen pre-S2 region into a different
 CC position of the FimH adhesin of type 1 fimbriae. Restriction site handles
 CC (BglII-sites) were introduced into the fimH gene, and the foreign
 CC epitopes are then inserted in-frame. In the selected positions the
 CC insertion of the epitopes did not significantly alter the adhesive
 CC function of the FimH protein. The expression of the chimeric proteins
 CC on the surface of fimbriae on bacterial hosts illustrated the possibility
 CC of using bacterial adhesins as general presenters of foreign antigens and
 CC epitopes. These chimeric genes may be used in the production of variant
 CC FimH adhesins which may be useful for targetting active compounds
 CC and microbial cells to locations comprising selected receptors to which
 CC the adhesins bind.
 CC
 SQ Sequence 23 AA;

Query Match 100.0%; Score 39; DB 16; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 EVPGSOH 7
 |||||

DB 6 evpgsqh 12
 RESULT 9
 ID AAP30265 standard; Protein; 26 AA.
 XX AAP30265;
 AC
 DT 21-APR-1992 (first entry)
 XX
 DE Sequence of amino acids 50-75 of the cholera toxin B1 subunit which
 DE carries an Arg at posns. 67 and 73.
 XX
 KW Cholera vaccine; therapy; E.coli infection; enterotoxin LT.
 XX
 OS Vibrio cholerae.
 OS
 XX EP95426-A.
 PN
 PD 30-NOV-1983.
 XX
 PF 26-MAY-1983; 83EP-0401052.
 XX
 PR 26-MAY-1982; 82FR-0009167.
 XX
 PA (CNRS) CNRS CENT NAT RECH SCI.
 PA (INSP) INST PASTEUR.
 XX
 PI Milhaud G, Raulais D, Rivallie P, Siffert O, Dodin A;
 XX
 DR WPI; 1983-834645/49.
 XX
 PT Cholera toxin B, sub-unit polypeptide(s) as vaccines and
 PT medicaments - effective against Escherichia coli and Vibrio
 PT cholerae infections, are prepd. by solid phase peptide synthesis
 PS
 PS Claim 7; Page 11; 13pp; French.
 XX
 CC The inventors claim cholera toxin B1 subunit sequences which carry
 CC Arg in posn. 35, 67 or 73 (AAP30264, AAP30265, AAP30600). Tests have
 CC shown that Arg in posns. 35, 67 and 73 play an important role in fixing
 CC these toxins to cell walls. The peptides are used in the treatment of,
 CC and vaccination against, cholera infections and animal and human
 CC infections due to E. coli (enterotoxin LT). The medicament may be
 CC administered by oral, intraperitoneal, sub-cutaneous or intravenous
 CC routes. For vaccines, pref. peptides having 15-30 AAs are injected
 CC without carriers. Unit dose when used as a medicament is 50-500mg as
 CC a vaccine 1-10mg of active cpd.
 CC
 SQ Sequence 26 AA;

Query Match 100.0%; Score 39; DB 4; Length 26;
 Best Local Similarity 100.0%; Pred. No. 0.65;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 EVPGSOH 7
 |||||

DB 2 evpgsqh 8
 RESULT 10
 ID AAP50439 standard; protein; 41 AA.
 XX AAP50439;
 AC
 DT 01-JAN-1980 (first entry)
 XX
 DE Network polymer which comprises a series of composite E. coli heat-
 DE labile toxin (LT)/heat-stable toxin (ST) polypeptide repeating units.

XX Heat-labile enterotoxin; heat-stable toxin; vaccine; ss.
 XX Synthetic.
 OS
 XX MO8502611-A.
 PN
 XX 20-JUN-1985.
 PD
 XX 12-DEC-1984; 84WO-US02030.
 PF
 XX 12-DEC-1983; 83US-0559469.
 PR
 XX (SCRI-) SCRIPPS CLINIC RES.
 PA
 XX Houghten RA;
 PI
 XX WPI; 1985-159230/26.
 DR
 XX
 PT New synthetic polypeptide(s) for part of E. coli enterotoxin(s) -
 useful for vaccination of mammals against the enterotoxin(s)
 PS
 XX Claim 8; Page 100; 120pp; English.
 PS
 XX The repeating units are bonded together by intramolecular
 CC interpolypeptide cystine bonds formed between oxidized Cys residues
 CC of the repeating units. This polypeptide may be used in the
 CC vaccination of mammals for protection against the enterotoxins. The
 CC composite polypeptide is made by solid phase synthesis or
 CC recombinant DNA techniques. X9 is Met or Lys, X12 is Ala or Glu,
 CC X39 is Asn or Tyr and X46 is Tyr or Asn. See also AAP50439-57.
 CC
 XX
 SQ Sequence 41 AA;
 OY
 1 EVPGSQH 7
 |||||
 Db 15 evpgsqh 21
 RESULT 11
 AAP50436
 ID AAP50436 standard; protein; 46 AA.
 AC
 XX AAP50436;
 AC
 XX 01-JAN-1980 (first entry)
 DT
 XX Network polymer which comprises a series of composite E. coli heat-
 DE labile toxin (LT)/heat-stable toxin (ST) polypeptide repeating units.
 DE
 XX Heat-labile enterotoxin; heat-stable toxin; vaccine; ss.
 KW
 XX Synthetic.
 OS
 XX MO8502611-A.
 PN
 XX 20-JUN-1985.
 PD
 XX 12-DEC-1984; 84WO-US02030.
 PF
 XX 12-DEC-1983; 83US-0559469.
 PR
 XX (SCRI-) SCRIPPS CLINIC RES.
 PA
 XX Houghten RA;
 PI
 XX WPI; 1985-159230/26.
 DR
 XX

PT New synthetic polypeptide(s) for part of E. coli enterotoxin(s) -
 useful for vaccination of mammals against the enterotoxin(s)
 PS
 XX Claim 8; Page 100; 120pp; English.
 PS
 XX The repeating units are bonded together by intramolecular
 CC interpolypeptide cystine bonds formed between oxidized Cys residues
 CC of the repeating units. This polypeptide may be used in the
 CC vaccination of mammals for protection against the enterotoxins. The
 CC composite polypeptide is made by solid phase synthesis or
 CC recombinant DNA techniques. X9 is Met or Lys, X12 is Ala or Glu,
 CC X39 is Asn or Tyr and X46 is Tyr or Asn. See also AAP50439-57.
 CC
 XX
 SQ Sequence 46 AA;
 OY
 1 EVPGSQH 7
 |||||
 Db 17 evpgsqh 23
 RESULT 12
 AAP30600
 ID AAP30600 standard; protein; 47 AA.
 AC
 XX AAP30600;
 AC
 XX 21-APR-1992 (first entry)
 DT
 XX Sequence of amino acids 350-75 of the cholera toxin B1 subunit which
 DE carries an Arg at posns. 35, 67 and 73.
 DE
 XX Cholera vaccine; therapy; E.coli infection; enterotoxin LT.
 KW
 XX Vibrio cholerae.
 OS
 XX EP95426-A.
 PN
 XX 30-NOV-1983.
 PD
 XX 26-MAY-1983; 83EP-0401052.
 PF
 XX 26-MAY-1982; 82FR-0009167.
 PR
 XX (CNRS) CNRS CENT NAT RECH SCT.
 PA (INSP) INST PASTEUR.
 XX
 PI Milhaud G, Raulais D, Rivaille P, Siffert O, Dodin A;
 PT Guyongruaz A, Delmas A;
 DR WPI; 1983-834645/49.
 XX
 PT Cholera toxin B, sub-unit polypeptide(s) as vaccines and
 PT medicaments - effective against Escherichia coli and Vibrio
 PT cholerae infections, are prepd. by solid phase peptide synthesis
 PS
 XX Claim 8; Page 11; 13pp; French.
 PS
 XX The inventors claim cholera toxin B1 subunit sequences which carry
 CC Arg in posn. 35, 67 or 73 (AAP30264, AAP30265, AAP30600). Tests have
 CC shown that Arg in posns. 35, 67 and 73 play an important role in fixing
 CC these toxins to cell walls. The peptides are used in the treatment of,
 CC and vaccination against, cholera infections and animal and human
 CC infections due to E. coli (enterotoxin LT). The medicament may be
 CC administered by oral, intraperitoneal, sub-cutaneous or intravenous
 CC routes. For vaccines, pref. peptides having 15-30 AAs are injected
 CC without carriers. Unit dose when used as a medicament is 50-500mg as
 CC a vaccine 1-10mg of active cpd.
 CC

SQ Sequence 47 AA;

Query Match 100.0%; Score 39; DB 4; Length 47;
 Best Local Similarity 100.0%; Pred. No. 1.2;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSOH 7

DB 23 evpgsqh 29

RESULT 13

ID AAR72545 standard; peptide; 93 AA.

AC AAR72545;

DT 28-NOV-1995 (first entry)

DE ADP-ribosylating toxin (verotoxin-1 B-subunit).

KM ADP-ribosylating toxin; pertussis holotoxin; B-subunit;

KM active site; E. coli heat labile toxin; verotoxin-1;

KM Bordetella pertussis vaccines.

OS Bacteria sp.

PN EP646599-A.

PD 05-APR-1995.

PF 23-AUG-1994; 94EP-0306219.

PR 24-AUG-1993; 93US-0110947.

PR 31-MAY-1994; 94US-0251121.

PA (CONN-) CONNAUGHT LAB LTD.

PA (UYAL-) UNIV ALBERTA.

PI Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;

PI Oomen R, Read RJ, Stein PE;

PS Disclosure; Fig 5; 54pp; English.

CC AAR72540-R72545 are structurally equivalent B-subunits from three
 CC ADP-ribosylating toxins, pertussis holotoxin (PT), E. coli heat
 CC labile toxin (LT), and verotoxin-1 (VT). The structural
 CC information obtd. from these comparisons was used to identify
 CC sites which contribute to PT's biological activity. By modifying
 CC these sites the claimed PT mutants of the invention were produced,
 CC they can be used in the development of vaccines against Bordetella
 CC pertussis infection.

SQ Sequence 93 AA;

Query Match 100.0%; Score 39; DB 16; Length 93;
 Best Local Similarity 100.0%; Pred. No. 2.4;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSOH 7

DB 41 evpgsqh 47

RESULT 14

ID AAY41816 standard; peptide; 93 AA.

AC AAY41816;

DT 08-DEC-1999 (first entry)

DE Escherichia coli verotoxin-1 B-subunit.

KM ADP-ribosylating toxin; PT; crystalline pertussis holotoxin; ETA; DT;
 KM three-dimensional structure; LT; immunoprotective; infection.

OS Escherichia coli.

PN US5965385-A.

PD 12-OCT-1999.

PF 06-JUN-1995; 95US-0467974.

PR 22-AUG-1994; 94US-0292968.

PR 24-AUG-1993; 93US-0110947.

PR 31-MAY-1994; 94US-0251121.

PA (CONN-) CONNAUGHT LAB LTD.

PA (UYAL-) UNIV ALBERTA.

PI Read RJ, Stein PE, Cockle SA, Loosmore S, Klein MH, Armstrong GD;
 PI Hazes B, Oomen RP;

DR WPI; 1999-579908/49.

PT New method for producing modified pertussis holotoxin -
 PT Example 3; Fig 5; 41pp; English.

CC A method has been developed of producing a modified pertussis holotoxin,
 CC involving analysis of the 3-dimensional form of the crystalline
 CC holotoxin. The pertussis holotoxin modification process comprises:
 CC (1) identification of at least one amino acid (aa) residue of the
 CC holotoxin for modification by analysing the 3-dimensional form of the
 CC crystalline holotoxin, in relation to known information of the protein
 CC structure and function; (2) effecting mutagenesis (by removing or
 CC replacing a nucleotide sequence encoding at least one (aa) of a tox
 CC operon; and (3) expressing mutant tox box in a Bordetella organism to
 CC produce the modified holotoxin. This method is used for modifying
 CC pertussis holotoxin, by studying its 3-dimensional crystalline
 CC structure. Modifying the holotoxin, alters its biological properties.
 CC By analysing the 3-dimensional crystalline structure of the pertussis
 CC holotoxin, functional (aa) which affect biological properties of the
 CC pertussis holotoxin can be identified. This can be used to predict (aa)
 CC which contribute to the toxicity of the holotoxin to produce
 CC immunoprotective, genetically-detoxified analogues of pertussis
 CC holotoxin. The present sequence represents an ADP-ribosylating toxin
 CC B-subunit peptide used in the exemplification of the present
 CC invention.

SQ Sequence 93 AA;

Query Match 100.0%; Score 39; DB 20; Length 93;
 Best Local Similarity 100.0%; Pred. No. 2.4;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSOH 7

DB 41 evpgsqh 47

RESULT 15

ID AAW95226 standard; peptide; 93 AA.

XX

AC AAW95226;
 XX
 DT 16-MAR-1999 (first entry)
 XX
 DE E. coli heat-labile toxin (LT) beta-subunit sequence.
 XX
 KW pertussis holotoxin; PT, modified; effector; toxicity; cell binding;
 KW enzyme; ADP-ribosylation; toxin; crystalline; X-ray crystallography;
 KW structural analysis; interacting site; mitogenicity; adjuvanticity;
 KW heat-labile; LT.
 XX
 OS Escherichia coli.
 XX
 PN US5856122-A.
 XX
 PD 05-JAN-1999.
 XX
 PF 22-AUG-1994; 94US-0292968.
 XX
 PR 22-AUG-1994; 94US-0292968.
 PR 24-AUG-1993; 93US-0110947.
 PR 31-MAY-1994; 94US-0251121.
 XX
 PA (UYAL-) UNIV ALBERTA.
 XX
 PI Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;
 PI Oomen RP, Read RJ, Stein PE;
 XX
 DR WPI: 1999-105104/09.
 XX
 PT Modifications to e.g. enzymatic activity, mitogenicity and cell
 PT binding of pertussis holotoxin - by identifying interaction sites of
 PT a molecule with crystalline toxin and modifying the identified site
 XX
 PS Example 3; Fig 5; 40pp; English.
 XX
 CC The invention relates to methods of preparing a pertussis holotoxin (PT)
 CC having a modified biological activity. One method comprises identifying
 CC at least 1 site in a PT that interacts with a molecule that is capable of
 CC forming a complex with the holotoxin and which molecule is an effector
 CC molecule which is an adenine nucleotide and which site contributes to
 CC toxicity, cell binding or enzymatic activity of PT. The functional
 CC interacting site(s) are identified by analysing the three dimensional
 CC structure of crystalline PT, determined by X-ray crystallography. The
 CC identified interacting site(s) are modified to alter toxicity, cell
 CC binding or enzyme activity of the PT. The methods can be used to alter a
 CC biological activity such as toxicity, enzymatic activity, mitogenicity,
 CC cell binding and adjuvanticity of the PT. The three-dimensional structure
 CC of PT have functional and/or structural resemblance to other bacterial
 CC toxins such as diphtheria toxin (DT), Pseudomonas exotoxin A (ETA), the
 CC heat-labile toxin of E. coli (LT) and verotoxin-1 (VT). The present
 CC sequence represents the beta-subunit of LT toxin.
 XX
 SQ Sequence 93 AA;

Query Match 100.0%; Score 39; DB 20; Length 93;
 Best Local Similarity 100.0%; Pred. No. 2.4;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSOH 7
 |||||
 Db 41 evpsqgh 47

Search completed: July 3, 2002, 08:54:56
 Job time: 286 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 08:52:05 ; Search time 34.88 Seconds
(without alignments)
4.902 Million cell updates/sec

Title: US-09-786-648-2

Perfect score: 39

Sequence: 1 EVPGSQH 7

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	100.0	93	2	US-08-292-968-26 Sequence 26, Appl
2	39	100.0	93	2	US-08-467-974-26 Sequence 26, Appl
3	39	100.0	93	2	US-08-467-536-26 Sequence 26, Appl
4	39	100.0	93	3	US-08-467-976-26 Sequence 26, Appl
5	39	100.0	93	4	US-09-082-514-26 Sequence 26, Appl
6	39	100.0	102	3	US-08-952-337-5 Sequence 5, Appl
7	39	100.0	102	3	US-08-952-337-6 Sequence 6, Appl
8	39	100.0	103	2	US-08-472-171-2 Sequence 2, Appl
9	39	100.0	103	2	US-08-894-526-2 Sequence 2, Appl
10	39	100.0	103	2	US-09-013-047-2 Sequence 2, Appl
11	39	100.0	103	4	US-09-374-597-2 Sequence 2, Appl
12	39	100.0	103	5	PCT-US95-13376-21 Sequence 21, Appl
13	39	100.0	103	5	PCT-US95-13376-21 Sequence 21, Appl
14	39	100.0	123	3	US-08-952-337-1 Sequence 1, Appl
15	39	100.0	123	3	US-08-952-337-2 Sequence 2, Appl
16	39	100.0	124	2	US-08-747-410-2 Sequence 2, Appl
17	39	100.0	371	1	US-08-829-026A-6 Sequence 6, Appl
18	33	84.6	124	1	US-08-449-045C-4 Sequence 4, Appl
19	33	84.6	124	6	US-08-435-605A-12 Sequence 12, Appl
20	33	84.6	124	6	US-08-435-605A-12 Sequence 12, Appl
21	31	79.5	93	4	US-09-069-023-15 Sequence 15, Appl
22	31	79.5	101	4	US-09-031-962D-16 Sequence 16, Appl
23	31	79.5	101	4	US-09-031-962D-16 Sequence 17, Appl
24	31	79.5	219	4	US-09-069-023-12 Sequence 12, Appl
25	31	79.5	414	1	US-08-255-471-9 Sequence 9, Appl
26	31	79.5	459	6	US-08-255-471-9 Patent No. 5194375
27	30	76.9	262	6	US-08-255-471-9 Patent No. 5194375

28	30	76.9	329	2	US-08-781-802-8	Sequence 8, Appl
29	30	76.9	329	4	US-08-694-078-8	Sequence 8, Appl
30	30	76.9	329	4	US-09-058-260-8	Sequence 8, Appl
31	30	76.9	459	6	US-08-714-070A-1	Sequence 1, Appl
32	30	76.9	775	2	US-08-045-806-2	Sequence 2, Appl
33	30	76.9	805	1	US-08-366-051B-2	Sequence 2, Appl
34	30	76.9	855	2	US-09-027-337-2	Sequence 2, Appl
35	30	76.9	855	2	US-08-540-406-6	Sequence 6, Appl
36	30	76.9	1285	3	US-08-556-055-6	Sequence 6, Appl
37	30	76.9	1285	3	US-08-954-668-6	Sequence 6, Appl
38	30	76.9	1285	5	PCT-US95-13233-6	Sequence 6, Appl
39	30	76.9	1286	4	US-09-268-140-3	Sequence 3, Appl
40	30	76.9	1286	4	US-08-460-900C-62	Sequence 62, Appl
41	30	76.9	1299	4	US-08-674-509B-48	Sequence 48, Appl
42	30	76.9	1299	4	US-08-954-698-48	Sequence 48, Appl
43	30	76.9	1810	5	PCT-US95-11684-4	Sequence 4, Appl
44	30	76.9	1810	5	PCT-US95-11684-4	Sequence 4, Appl
45	29	74.4	346	2	US-08-602-359A-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1
US-08-292-968-26
Sequence 26, Application US/08292968
Patent No. 5856122
GENERAL INFORMATION:
APPLICANT: READ, Randy J.
APPLICANT: STEIN, Penelope E.
APPLICANT: COCKLE, Stephen A.
APPLICANT: OOMEN, Raymond P.
APPLICANT: LOOSMORE, Sheena
APPLICANT: KLEIN, Michel H.
APPLICANT: ARMSTRONG, Glen D.
APPLICANT: HAZES, Bart
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Slim & McBurney
STREET: Suite 701, 330 University Avenue
City: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/292,968
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/110,947
FILING DATE: 24-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/251,121
FILING DATE: 31-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-388
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEO ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-292-968-26

Query Match 100.0%; Score 39; DB 2; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7
|||||
DB 41 EYVGSQH 47

US-08-467-974-26

; Sequence 26, Application US/08467974
; Patent No. 5965385

; GENERAL INFORMATION:

; APPLICANT: READ, Randy J.

; APPLICANT: STEIN, Penelope E.

; APPLICANT: COCKLE, Stephen A.

; APPLICANT: COOMEN, Raymond P.

; APPLICANT: LOOSMORE, Sheena

; APPLICANT: KLEIN, Michel H.

; APPLICANT: ARMSTRONG, Glen D.

; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN

; NUMBER OF SEQUENCES: 46

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: Suite 701, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/467,974

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/467,536

; FILING DATE: 06-JUN-1995

; APPLICATION NUMBER: US 08/292,968

; FILING DATE: 22-AUG-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/251,121

; FILING DATE: 31-MAY-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/110,947

; FILING DATE: 24-AUG-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: STEWART, Michael I.

; REGISTRATION NUMBER: 24,973

; REFERENCE/DOCKET NUMBER: 1038-454 MIS:vg

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1163

; INFORMATION FOR SEQ ID NO: 26:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 93 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-467-974-26

Query Match 100.0%; Score 39; DB 2; Length 93;

Best Local Similarity 100.0%; Pred. No. 0.67;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7
|||||
DB 41 EYVGSQH 47

US-08-467-536-26

; Sequence 26, Application US/08467536
; Patent No. 5977304

; GENERAL INFORMATION:

; APPLICANT: READ, Randy J.

; APPLICANT: STEIN, Penelope E.

; APPLICANT: COCKLE, Stephen A.

; APPLICANT: COOMEN, Raymond P.

; APPLICANT: LOOSMORE, Sheena

; APPLICANT: KLEIN, Michel H.

; APPLICANT: ARMSTRONG, Glen D.

; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN

; NUMBER OF SEQUENCES: 46

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: Suite 701, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/467,536

; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/292,968

; FILING DATE: 22-AUG-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/251,121

; FILING DATE: 31-MAY-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/110,947

; FILING DATE: 24-AUG-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: STEWART, Michael I.

; REGISTRATION NUMBER: 24,973

; REFERENCE/DOCKET NUMBER: 1038-455 MIS:vg

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1163

; INFORMATION FOR SEQ ID NO: 26:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 93 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-467-536-26

Query Match 100.0%; Score 39; DB 2; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7
|||||
DB 41 EYVGSQH 47

US-08-467-976-26

Query Match 100.0%; Score 39; DB 2; Length 93;

Best Local Similarity 100.0%; Pred. No. 0.67;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Sequence 26, Application US/08467976
Patent No. 6018022
GENERAL INFORMATION:
APPLICANT: READ, Randy J.
APPLICANT: STEIN, Penelope E.
APPLICANT: COCKLE, Stephen A.
APPLICANT: OOMEN, Raymond P.
APPLICANT: LOOSMORE, Sheena
APPLICANT: KLEIN, Michel H.
APPLICANT: ARMSTRONG, Glen D.
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,976
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/292,968
FILING DATE: 22-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/251,121
FILING DATE: 31-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/110,947
FILING DATE: 24-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-453 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-467-976-26

Query Match 100.0%; Score 39; DB 3; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSOH 7
|||||
DB 41 EYPGSOH 47

RESULT 5
US-09-082-514-26
Sequence 26, Application US/09082514
Patent No. 6168928
GENERAL INFORMATION:
APPLICANT: READ, Randy J.
APPLICANT: STEIN, Penelope E.
APPLICANT: COCKLE, Stephen A.
APPLICANT: OOMEN, Raymond P.
APPLICANT: KLEIN, Michel H.

APPLICANT: ARMSTRONG, Glen D.
APPLICANT: HAZES, Bart
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/082,514
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/292,968
FILING DATE: 24-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-810
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-082-514-26

Query Match 100.0%; Score 39; DB 4; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPGSOH 7
|||||
DB 41 EYPGSOH 47

RESULT 6
US-08-952-337-5
Sequence 5, Application US/08952337
Patent No. 6019773
GENERAL INFORMATION:
APPLICANT: Holmgren, Jan
APPLICANT: Iebens, Michael R.
TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS
FILE REFERENCE: 3846/0D758
CURRENT APPLICATION NUMBER: US/08/952,337
CURRENT FILING DATE: 1998-01-05
EARLIER APPLICATION NUMBER: PCT/SE96/00570
EARLIER FILING DATE: 1996-05-02
EARLIER APPLICATION NUMBER: SE 9501682-0
EARLIER FILING DATE: 1995-05-05
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5
LENGTH: 102
TYPE: PRT
ORGANISM: Vibrio cholerae
US-08-952-337-5

Query Match 100.0%; Score 39; DB 3; Length 102;
Best Local Similarity 100.0%; Pred. No. 0.73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYPSGSH 7
|||||||
DB 50 EYPSGSH 56

RESULT 7
US-08-952-337-6

; Sequence 6, Application US/08952337
; Patent No. 6019973
; GENERAL INFORMATION:
; APPLICANT: Holmgren, Jan
; APPLICANT: Lebens, Michael R.
; TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
; TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS
; FILE REFERENCE: 3846/OD758
; CURRENT FILING DATE: 1998-01-05
; EARLIER APPLICATION NUMBER: PCT/SE96/00570
; EARLIER FILING DATE: 1996-05-02
; EARLIER APPLICATION NUMBER: SE 9501682-0
; EARLIER FILING DATE: 1995-05-05
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-952-337-6

Query Match 100.0%; Score 39; DB 3; Length 102;
Best Local Similarity 100.0%; Pred. No. 0.73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYPSGSH 7
|||||||
DB 50 EYPSGSH 56

RESULT 8
US-08-472-171-2

; Sequence 2, Application US/08472171
; Patent No. 5932714
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yacoub, Reza K.
; APPLICANT: Zealey, Gavin R.
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Expression Of Gene Products From
; TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 330 University Avenue, Suite 701
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/393,334

; FILING DATE: 23-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-507 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 416-595-1155
; TELEFAX: 416-595-1163
; TELEX: 065-24567 Simbas
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-472-171-2

Query Match 100.0%; Score 39; DB 2; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYPSGSH 7
|||||||
DB 51 EYPSGSH 57

RESULT 9
US-08-894-526-2

; Sequence 2, Application US/08894526
; Patent No. 5942418
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Yacoub, Reza K
; APPLICANT: Zealey, Gavin R
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: EXPRESSION OF GENE PRODUCTS FROM
; TITLE OF INVENTION: GENETICALLY MANIPULATED STRAINS OF BORDETELLA
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,526
; FILING DATE: 01-DEC-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-724 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-894-526-2

Query Match 100.0%; Score 39; DB 2; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.74;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
Db 51 EVPGSOH 57

RESULT 10

US-09-013-047-2
Sequence 2, Application US/09013047
Patent No. 5998168
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M.
APPLICANT: Yacoub, Reza K.
APPLICANT: Zealey, Gavin R.
APPLICANT: Klein, Michael H.
TITLE OF INVENTION: Expression Of Gene Products From
TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & Mcburney
STREET: 330 University Avenue, 6th Floor
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/013,047
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/472,171
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/393,334
FILING DATE: 23-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-507 MIS:Vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: 416-595-1155
TELEFAX: 416-595-1163
TELEX: 065-24567 Simbas
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-013-047-2

Query Match 100.0%; Score 39; DB 2; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
Db 51 EVPGSOH 57

RESULT 11
US-09-374-597-2
Sequence 2, Application US/09374597
Patent No. 6140082
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M.

APPLICANT: Yacoub, Reza K.
APPLICANT: Zealey, Gavin R.
APPLICANT: Klein, Michael H.
TITLE OF INVENTION: Expression Of Gene Products From
TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & Mcburney
STREET: 330 University Avenue, 6th Floor
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/374,597
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/393,334
FILING DATE: FEBRUARY 23, 1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24973
REFERENCE/DOCKET NUMBER: 1038-964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 416-595-1155
TELEFAX: 416-595-1163
TELEX: 065-24567 Simbas
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-374-597-2

Query Match 100.0%; Score 39; DB 4; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
Db 51 EVPGSOH 57

RESULT 12
US-09-191-852-21
Sequence 21, Application US/09191852
Patent No. 6194560
GENERAL INFORMATION:
APPLICANT: Charles J. Arntzen, Hugh S. Mason, and Tariq A. Haq
TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski L.L.P.
STREET: 1301 McKinney, Suite 5100
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77010
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/191,852
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13376
FILING DATE: 24-OCT-1995
APPLICATION NUMBER: 08/817,906
FILING DATE: 04-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Fox, David L.
REGISTRATION NUMBER: 40,612
REFERENCE/DOCKET NUMBER: P015900S1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-651-5151
TELEFAX: 713-651-5246
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-09-191-852-21

Query Match 100.0%; Score 39; DB 4; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
|||||||
DB 51 EVPGSOH 57

RESULT 13
PCT-US95-13376-21
Sequence 21, Application PC/TUS9513376
GENERAL INFORMATION:
APPLICANT: The Texas A&M University System
APPLICANT: 310 Wisenbaker
APPLICANT: College Station, Texas 77843-3369
TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pravel, Hewitt, Kimball & Krieger
STREET: 1177 West Loop South, 10th Floor
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77027-9095
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13376
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/328,716
FILING DATE: 24-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Jones, John W.
REGISTRATION NUMBER: 31,380
REFERENCE/DOCKET NUMBER: 36170/3P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-850-0909
TELEFAX: 713-850-0165
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear
PCT-US95-13376-21

Query Match 100.0%; Score 39; DB 5; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
|||||||
DB 51 EVPGSOH 57

RESULT 14
US-08-952-337-1
Sequence 1, Application US/08952337
Patent No. 6019973
GENERAL INFORMATION:
APPLICANT: Holmgren, Jan
APPLICANT: Lebens, Michael R.
TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS
FILE REFERENCE: 3846/OD758
CURRENT APPLICATION NUMBER: US/08/952,337
CURRENT FILING DATE: 1998-01-05
EARLIER APPLICATION NUMBER: PCT/SE96/00570
EARLIER FILING DATE: 1996-05-02
EARLIER APPLICATION NUMBER: SE 9501682-0
EARLIER FILING DATE: 1995-05-05
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 123
TYPE: PRT
ORGANISM: Vibrio cholerae
US-08-952-337-1

Query Match 100.0%; Score 39; DB 3; Length 123;
Best Local Similarity 100.0%; Pred. No. 0.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
|||||||
DB 71 EVPGSOH 77

RESULT 15
US-08-952-337-2
Sequence 2, Application US/08952337
Patent No. 6019973
GENERAL INFORMATION:
APPLICANT: Holmgren, Jan
APPLICANT: Lebens, Michael R.
TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS
FILE REFERENCE: 3846/OD758
CURRENT APPLICATION NUMBER: US/08/952,337
CURRENT FILING DATE: 1998-01-05
EARLIER APPLICATION NUMBER: PCT/SE96/00570
EARLIER FILING DATE: 1996-05-02
EARLIER APPLICATION NUMBER: SE 9501682-0
EARLIER FILING DATE: 1995-05-05
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 123
TYPE: PRT
ORGANISM: Escherichia coli
US-08-952-337-2

Query Match 100.0%; Score 39; DB 3; Length 123;
Best Local Similarity 100.0%; Pred. No. 0.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	EVPGSQH	7
Db	71	EVPGSQH	77

Search completed: July 3, 2002, 08:55:38
Job time: 213 sec

THIS PAGE BLANK (USPTO)

Gencore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 08:53:00 ; Search time 401.04 Seconds
(without alignments)
6.144 Million cell updates/sec

Title: US-09-786-648-2
Perfect score: 39
Sequence: 1 EVPGSQH 7

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues
Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending Patents AA Main:*

1: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US10_COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US10_COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US10_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	100.0	7	21	US-09-786-648-2
2	39	100.0	12	21	US-09-786-648-3
3	39	100.0	15	11	US-08-732-371-1
4	39	100.0	15	11	US-08-732-371A-1
5	39	100.0	21	21	US-09-786-648-4
6	39	100.0	21	21	US-09-786-648-5
7	39	100.0	93	5	US-08-110-947-10

8	39	100.0	93	5	US-08-110-947A-26	Sequence 26, Appl
9	39	100.0	93	6	US-08-251-121-26	Sequence 26, Appl
10	39	100.0	103	7	US-08-393-334-2	Sequence 2, Appl
11	39	100.0	103	11	US-08-782-832-15	Sequence 15, Appl
12	39	100.0	103	12	US-08-817-906-21	Sequence 21, Appl
13	39	100.0	103	22	US-09-836-433-14	Sequence 14, Appl
14	39	100.0	116	22	US-09-836-433-20	Sequence 20, Appl
15	39	100.0	119	22	US-09-836-433-22	Sequence 22, Appl
16	39	100.0	124	1	PCT-US99-30747-55	Sequence 55, Appl
17	39	100.0	124	1	PCT-US99-30747-57	Sequence 55, Appl
18	39	100.0	124	18	US-09-470-124-55	Sequence 55, Appl
19	39	100.0	124	18	US-09-470-124-57	Sequence 57, Appl
20	39	100.0	138	13	US-08-914-479-2	Sequence 2, Appl
21	39	100.0	138	13	US-08-914-479A-2	Sequence 2, Appl
22	39	100.0	133	21	US-09-756-983-15	Sequence 15, Appl
23	39	100.0	131	21	US-09-756-983-18	Sequence 18, Appl
24	39	100.0	364	21	US-09-756-983-22	Sequence 22, Appl
25	39	100.0	371	5	US-08-150-305A-3	Sequence 3, Appl
26	39	100.0	371	11	US-08-784-218-6	Sequence 6, Appl
27	39	100.0	371	12	US-08-829-026-5	Sequence 5, Appl
28	39	89.7	142	26	US-60-361-742-1547	Sequence 1547, Ap
29	39	87.2	41	20	US-09-688-051-2967	Sequence 2967, Ap
30	34	87.2	84	20	US-09-617-682A-5651	Sequence 5651, Ap
31	34	87.2	124	18	US-09-417-507-29272	Sequence 29272, A
32	34	87.2	304	1	PCT-US01-08631-32859	Sequence 32859, A
33	34	87.2	346	1	PCT-US01-08631-32459	Sequence 32459, A
34	34	87.2	578	1	PCT-US01-08631-32551	Sequence 32551, A
35	34	87.2	847	1	PCT-US01-08631-54035	Sequence 54035, A
36	33	84.6	51	22	US-09-866-066-35	Sequence 35, Appl
37	33	84.6	124	21	US-09-760-234-7	Sequence 7, Appl
38	33	84.6	131	26	US-60-171-481-1502	Sequence 1502, Ap
39	33	84.6	318	1	PCT-US99-17130-318	Sequence 318, App
40	33	84.6	318	18	US-09-489-847-328	Sequence 328, App
41	33	84.6	379	18	US-09-402-532-1	Sequence 1, Appl
42	33	84.6	379	18	US-09-402-532-3	Sequence 3, Appl
43	33	84.6	382	1	PCT-US01-08582-3	Sequence 3, Appl
44	33	84.6	382	1	PCT-US01-08582-4	Sequence 4, Appl
45	33	84.6	382	18	US-09-402-532-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1

US-09-786-648-2

Sequence 2, Application US/09786648

GENERAL INFORMATION:

APPLICANT: Williams, Neil Andrew

TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vacci

FILE REFERENCE: 7438

CURRENT APPLICATION NUMBER: US/09/786, 648

PRIOR APPLICATION NUMBER: 2001-03-07

PRIOR FILING DATE: 1999-09-07

NUMBER OF SEQ ID NOS: 6

SOFTWARE: MS DOS

SEQ ID NO 2

LENGTH: 7

TYPE: PRT

ORGANISM: E. coli

FEATURE:

LOCATION: 51...57

OTHER INFORMATION: Isolated or synthetic E. coli

OTHER INFORMATION: human variant E. coli

US-09-786-648-2

Query Match 100.0%; Score 39; DB 21; Length 7;
Best Local Similarity 100.0%; Pred. No. 3.2e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7
DB 1 EYVGSQH 7

RESULT 2

US-09-786-648-3
Sequence 3, Application US/09786648
GENERAL INFORMATION:
APPLICANT: Williams, Neil Andrew
APPLICANT: Hirst, Timothy Raymond
TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccine
TITLE OF INVENTION: Adjuvants
FILE REFERENCE: 7438
CURRENT APPLICATION NUMBER: US/09/786,648
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: PCT/GB99/02970
PRIOR FILING DATE: 1999-09-07
NUMBER OF SEQ ID NOS: 6
SOFTWARE: MS DOS
SEQ ID NO 3
LENGTH: 12
TYPE: PRT
ORGANISM: E. coli
FEATURE:
LOCATION: 50...61
OTHER INFORMATION: Isolated or synthetic EtXB beta4-alpha2 loop fragment derivable f
US-09-786-648-3

Query Match 100.0%; Score 39; DB 21; Length 12;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7
DB 2 EYVGSQH 8

RESULT 3

US-08-732-371-1
Sequence 1, Application US/08732371
GENERAL INFORMATION:
APPLICANT: MIRELMAN, David
APPLICANT: MARKS, Robert S.
APPLICANT: SELA, Michael
TITLE OF INVENTION: VACCINES FOR ORAL IMMUNIZATION AGAINST
TITLE OF INVENTION: INFECTING AGENTS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/732,371
FILING DATE: 09-JAN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 109519
FILING DATE: 03-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: MIRELMAN-3

TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-732-371-1

Query Match 100.0%; Score 39; DB 11; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7
DB 2 EYVGSQH 8

RESULT 4

US-08-732-371A-1
Sequence 1, Application US/08732371A
GENERAL INFORMATION:
APPLICANT: MIRELMAN, David
APPLICANT: MARKS, Robert S.
APPLICANT: SELA, Michael
TITLE OF INVENTION: VACCINES FOR ORAL IMMUNIZATION AGAINST
TITLE OF INVENTION: INFECTING AGENTS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/732,371A
FILING DATE: 09-JAN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 109519
FILING DATE: 03-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: MIRELMAN-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-732-371A-1

Query Match 100.0%; Score 39; DB 11; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7
DB 2 EYVGSQH 8

RESULT 5
US-09-786-648-4
; Sequence 4, Application US/09786648
; GENERAL INFORMATION:
; APPLICANT: Williams, Neil Andrew
; APPLICANT: Hirst, Timothy Raymond
; TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccine
; FILE REFERENCE: 7438
; CURRENT APPLICATION NUMBER: US/09/786,648
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: PCT/GB99/02970
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: MS DOS
; SEQ ID NO 4
; LENGTH: 21
; TYPE: PRT
; ORGANISM: E. coli
; FEATURE:
; LOCATION: 45...65
; OTHER INFORMATION: isolated or synthetic EtXB beta4-alpha2 loop fragment derivable
; OTHER INFORMATION: human variant E. coli
US-09-786-648-4

Query Match 100.0%; Score 39; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
|||||
DB 7 EVPGSOH 13

RESULT 6
US-09-786-648-5
; Sequence 5, Application US/09786648
; GENERAL INFORMATION:
; APPLICANT: Williams, Neil Andrew
; APPLICANT: Hirst, Timothy Raymond
; TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccine
; FILE REFERENCE: 7438
; CURRENT APPLICATION NUMBER: US/09/786,648
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: PCT/GB99/02970
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: MS DOS
; SEQ ID NO 5
; LENGTH: 21
; TYPE: PRT
; ORGANISM: E. coli
; FEATURE:
; LOCATION: 45...65
; OTHER INFORMATION: isolated or synthetic EtXB beta4-alpha2 loop fragment derivable
; OTHER INFORMATION: porcine E. coli
US-09-786-648-5

Query Match 100.0%; Score 39; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
|||||
DB 7 EVPGSOH 13

RESULT 7

US-08-110-947-10
; Sequence 10, Application US/08110947
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J
; APPLICANT: STEIN, Penelope E
; APPLICANT: COCKLE, Stephen A
; APPLICANT: OOMEN, Raymond P
; APPLICANT: KLEIN, Michel H
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: Suite 1203 Crystal Plaza I, 2001 Jefferson
; STREET: Davis Hwy.
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,947
; FILING DATE: 24-AUG-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FALLOW, Charles W
; REGISTRATION NUMBER: 28,946
; REFERENCE/DOCKET NUMBER: 1038-303 MTS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; TELEX: 89-9456 LUKPAT
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-110-947-10

Query Match 100.0%; Score 39; DB 5; Length 93;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
|||||
DB 41 EVPGSOH 47

RESULT 8
US-08-110-947A-26
; Sequence 26, Application US/08110947A
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J
; APPLICANT: STEIN, Penelope E
; APPLICANT: COCKLE, Stephen A
; APPLICANT: OOMEN, Raymond P
; APPLICANT: KLEIN, Michel H
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: Suite 1203 Crystal Plaza I, 2001 Jefferson
; STREET: Davis Highway
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/110,947A
FILING DATE: 24-AUG-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: FALLOW, Charles W
REGISTRATION NUMBER: 28,946
REFERENCE/DOCKET NUMBER: 1038-303 MIS-1b
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
TELEX: 89-9456 LUKPAR
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-110-947A-26

Query Match 100.0%; Score 39; DB 5; Length 93;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
| | | | | | |
DB 41 EVPGSOH 47

RESULT 9

US-08-251-121-26
Sequence 26, Application US/08251121

GENERAL INFORMATION:

APPLICANT: READ, Randy J.
APPLICANT: STEIN, Penelope E.
APPLICANT: COCKLE, Stephen A.
APPLICANT: COOMEN, Raymond P.
APPLICANT: KLEIN, Michael H.
APPLICANT: ARMSTRONG, Glen D.

APPLICANT: HAZES, Bart

TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN

NUMBER OF SEQUENCES: 46

CORRESPONDENCE ADDRESS:

ADDRESSEE: Slim & McBurney

STREET: Suite 701, 330 University Avenue

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/251,121

FILING DATE: 31-MAY-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/110,947

FILING DATE: 24-AUG-1993

ATTORNEY/AGENT INFORMATION:

NAME: STEWART, Michael I.

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-335

TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-251-121-26

Query Match 100.0%; Score 39; DB 6; Length 93;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
| | | | | | |
DB 41 EVPGSOH 47

RESULT 10

US-08-393-334-2
Sequence 2, Application US/08393334

GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena M.

APPLICANT: Yacoub, Reza K.

APPLICANT: Zealey, Gavin R.

TITLE OF INVENTION: Expression Of Gene Products From

NUMBER OF SEQUENCES: Genetically Manipulated Strains Of Bordetella

CORRESPONDENCE ADDRESS: 56

ADDRESSEE: Slim & McBurney

STREET: 330 University Avenue, Suite 701

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/393,334

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I

REGISTRATION NUMBER: 24973

REFERENCE/DOCKET NUMBER: 1038-417

TELECOMMUNICATION INFORMATION:

TELEPHONE: 416-595-1163

TELEFAX: 065-24567 Simbas

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 103 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-393-334-2

Query Match 100.0%; Score 39; DB 7; Length 103;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
| | | | | | |
DB 51 EVPGSOH 57

RESULT 11
US-08-782-832-15

Sequence 15, Application US/08782832
GENERAL INFORMATION:
APPLICANT: Arnzen, Charles J.
APPLICANT: Mason, Hugh S.
TITLE OF INVENTION: PRODUCTION OF AN ORALLY IMMUNOGENIC
TITLE OF INVENTION: BACTERIAL PROTEIN IN TRANSGENIC PLANTS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pravel, Hewitt, Kimball & Krieger
STREET: 1177 West Loop South, 10th floor
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77027-9095
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/782,832
FILING DATE: 13-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/328,716
FILING DATE: 24-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ketteberger, Denise M.
REGISTRATION NUMBER: 33,924
REFERENCE/DOCKET NUMBER: 36170/3
TELEPHONE: 713-850-0909
TELEFAX: 713-850-0165
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-782-832-15

Query Match 100.0%; Score 39; DB 11; Length 103;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSOH 7
|||||||
Db 51 EVPGSOH 57

RESULT 12
US-08-817-906-21
Sequence 21, Application US/08817906
GENERAL INFORMATION:
APPLICANT: Charles J. Arnzen, Hugh S. Mason, John D. Clements,
APPLICANT: Haq, Tariq A.
TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski L.L.P.
STREET: 1301 McKinney, Suite 5100
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77010
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/817,906

FILING DATE: 08/04/97
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13376
FILING DATE: 24-OCT-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fox, David L.
REGISTRATION NUMBER: 40,612
REFERENCE/DOCKET NUMBER: 36170/3P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-651-5151
TELEFAX: 713-651-5246
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-817-906-21

Query Match 100.0%; Score 39; DB 12; Length 103;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSOH 7
|||||||
Db 51 EVPGSOH 57

RESULT 13
US-09-836-433-14
Sequence 14, Application US/09836433
GENERAL INFORMATION:
APPLICANT: Yuki, Yoshikazu
APPLICANT: Uda, Shigezo
TITLE OF INVENTION: HYBRID PROTEINS FOR AUTOIMMUNE DISEASE
FILE REFERENCE: Not Assigned
CURRENT APPLICATION NUMBER: US/09/836,433
CURRENT FILING DATE: 2001-04-16
NUMBER OF SEQ ID NOS: 49
SOFTWARE: Patentin version 3.0
SEQ ID NO 14
LENGTH: 103
TYPE: PRT
ORGANISM: Vibrio cholerae
US-09-836-433-14

Query Match 100.0%; Score 39; DB 22; Length 103;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSOH 7
|||||||
Db 51 EVPGSOH 57

RESULT 14
US-09-836-433-20
Sequence 20, Application US/09836433
GENERAL INFORMATION:
APPLICANT: Yuki, Yoshikazu
APPLICANT: Uda, Shigezo
TITLE OF INVENTION: HYBRID PROTEINS FOR AUTOIMMUNE DISEASE
FILE REFERENCE: Not Assigned
CURRENT APPLICATION NUMBER: US/09/836,433
CURRENT FILING DATE: 2001-04-16
NUMBER OF SEQ ID NOS: 49
SOFTWARE: Patentin version 3.0
SEQ ID NO 20
LENGTH: 116
TYPE: PRT
ORGANISM: synthetic construct

US-09-836-433-20

Query Match 100.0%; Score 39; DB 22; Length 116;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7
 |||||
 DB 51 EYVGSQH 57

RESULT 15

US-09-836-433-22
 ; Sequence 22, Application US/09836433
 ; GENERAL INFORMATION:
 ; APPLICANT: Yuki, Yoshikazu
 ; APPLICANT: Udaoka, Shigezo
 ; TITLE OF INVENTION: HYBRID PROTEINS FOR AUTOIMMUNE DISEASE
 ; FILE REFERENCE: Not Assigned
 ; CURRENT APPLICATION NUMBER: US/09/836,433
 ; CURRENT FILING DATE: 2001-04-16
 ; NUMBER OF SEQ ID NOS: 49
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 22
 ; LENGTH: 119
 ; TYPE: PRT
 ; ORGANISM: synthetic construct
 US-09-836-433-22

Query Match 100.0%; Score 39; DB 22; Length 119;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7
 |||||
 DB 51 EYVGSQH 57

Search completed: July 3, 2002, 09:03:20
 Job time: 620 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 08:53:20 ; Search time 46.52 Seconds
(Without alignments)
15.224 Million cell updates/sec

Title: US-09-786-648-2

Perfect score: 39

Sequence: 1 EVPGSOH 7

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 308740 segs, 101176262 residues

Total number of hits satisfying chosen parameters: 308740

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database :

Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	100.0	103	6	US-10-110-364-8 Sequence 8, Appl
2	39	100.0	103	6	US-10-110-364-10 Sequence 10, Appl
3	39	100.0	103	6	US-10-110-364-13 Sequence 13, Appl
4	39	100.0	103	6	US-10-110-364-22 Sequence 22, Appl
5	39	100.0	105	6	US-10-110-364-12 Sequence 12, Appl
6	39	100.0	123	6	US-10-110-364-17 Sequence 17, Appl
7	39	100.0	123	6	US-10-110-364-20 Sequence 20, Appl
8	39	100.0	124	6	US-10-110-364-2 Sequence 2, Appl
9	39	100.0	124	6	US-10-110-364-6 Sequence 6, Appl
10	39	100.0	124	6	US-10-110-364-11 Sequence 11, Appl
11	39	100.0	124	6	US-10-110-364-16 Sequence 16, Appl
12	39	100.0	124	6	US-10-110-364-18 Sequence 18, Appl
13	39	100.0	124	6	US-10-110-364-19 Sequence 19, Appl
14	39	100.0	124	6	US-10-110-364-21 Sequence 21, Appl
15	39	100.0	124	6	US-10-110-364-23 Sequence 23, Appl
16	39	100.0	125	6	US-10-110-364-15 Sequence 15, Appl
17	39	100.0	138	6	US-10-141-627-2 Sequence 2, Appl
18	36	92.3	103	6	US-10-110-364-7 Sequence 7, Appl
19	33	84.6	103	6	US-10-110-364-5 Sequence 5, Appl
20	33	84.6	124	6	US-10-110-364-4 Sequence 4, Appl
21	33	84.6	124	6	US-10-110-364-9 Sequence 9, Appl
22	33	84.6	382	5	US-09-809-033A-3 Sequence 9, Appl
23	33	84.6	382	5	US-09-809-033A-4 Sequence 4, Appl
24	33	84.6	412	5	US-09-978-403A-157 Sequence 157, App
25	33	84.6	412	5	US-09-978-544A-157 Sequence 157, App
26	33	84.6	412	5	US-09-978-681A-157 Sequence 157, App

27	33	84.6	412	5	US-09-978-757A-157 Sequence 157, App
28	33	84.6	412	5	US-09-978-564A-157 Sequence 157, App
29	33	84.6	412	5	US-09-999-831A-157 Sequence 157, App
30	33	84.6	412	5	US-09-999-829A-157 Sequence 157, App
31	33	84.6	412	5	US-09-978-375A-157 Sequence 157, App
32	33	84.6	412	5	US-09-978-423A-157 Sequence 157, App
33	33	84.6	412	6	US-10-013-921A-157 Sequence 157, App
34	33	84.6	412	6	US-10-013-929A-157 Sequence 157, App
35	33	84.6	412	6	US-10-013-918A-157 Sequence 157, App
36	33	84.6	412	6	US-10-017-082A-157 Sequence 157, App
37	33	84.6	412	6	US-10-017-085A-157 Sequence 157, App
38	33	84.6	412	6	US-10-013-916A-157 Sequence 157, App
39	33	84.6	412	6	US-10-017-086A-157 Sequence 157, App
40	33	84.6	412	6	US-10-013-925A-157 Sequence 157, App
41	33	84.6	412	6	US-10-017-081A-157 Sequence 157, App
42	33	84.6	412	6	US-10-016-177A-157 Sequence 157, App
43	33	84.6	412	6	US-10-017-084A-157 Sequence 157, App
44	33	84.6	412	6	US-10-013-923A-157 Sequence 157, App
45	33	84.6	473	6	US-10-155-881-26127 Sequence 26127, A

ALIGNMENTS

```
RESULT 1
US-10-110-364-8
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110.364
; PRIOR FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)..(103)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 998409
US-10-110-364-8

Query Match 100.0%; Score 39; DB 6; Length 103;
Best Local Similarity 100.0%; Pred. NO. 0.79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0./

OY 1 EVPGSOH 7
Db 51 EVPGSOH 57

RESULT 2
US-10-110-364-10
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110.364
```

```
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(103)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 1421511
US-10-110-364-10
```

```
Query Match          100.0%; Score 39; DB 6; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 EYPGSOH 7
    |||||
Db 51 EYPGSOH 57
```

```
RESULT 3
US-10-110-364-13
; Sequence 13, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(103)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 14215235
; OTHER INFORMATION: (Ogawa 41 R35D).
US-10-110-364-13
```

```
Query Match          100.0%; Score 39; DB 6; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 EYPGSOH 7
    |||||
Db 51 EYPGSOH 57
```

```
RESULT 4
US-10-110-364-22
; Sequence 22, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
```

```
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(103)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 494265.
US-10-110-364-22
```

```
Query Match          100.0%; Score 39; DB 6; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0,
```

```
OY 1 EYPGSOH 7
    |||||
Db 51 EYPGSOH 57
```

```
RESULT 5
US-10-110-364-12
; Sequence 12, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(105)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 2781121 (Ogawa
US-10-110-364-12
```

```
Query Match          100.0%; Score 39; DB 6; Length 105;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0,
```

```
OY 1 EYPGSOH 7
    |||||
Db 52 EYPGSOH 58
```

```
RESULT 6
US-10-110-364-17
; Sequence 17, Application US/10110364
```



```

; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewall, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; PRIOR FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(123)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 1395122.
US-10-110-364-17

```

```

Query Match          100.0%; Score 39; DB 6; Length 123;
Best Local Similarity 100.0%; Pred. No. 0.96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 EVPGSOH 7
DB 72 EVPGSOH 78

```

```

RESULT 7
US-10-110-364-20
; Sequence 20, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewall, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; PRIOR FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(123)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 223254.
US-10-110-364-20

```

```

Query Match          100.0%; Score 39; DB 6; Length 123;
Best Local Similarity 100.0%; Pred. No. 0.96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 EVPGSOH 7
DB 71 EVPGSOH 77

```

```

RESULT 8
US-10-110-364-2
; Sequence 2, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewall, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; PRIOR FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Vibrio cholera
; US-10-110-364-2

```

```

Query Match          100.0%; Score 39; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 EVPGSOH 7
DB 72 EVPGSOH 78

```

```

RESULT 9
US-10-110-364-6
; Sequence 6, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewall, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; PRIOR FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 808900.
US-10-110-364-6

```

```

Query Match          100.0%; Score 39; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 EVPGSOH 7
DB 72 EVPGSOH 78

```

RESULT 10

US-10-110-364-11
; Sequence 11, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparenta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: US/10/110.364
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 48890 (classic
; OTHER INFORMATION: 569b).
US-10-110-364-11

Query Match 100.0%; Score 39; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 EVPGSQH 7
1111111
Db 72 EVPGSQH 78

RESULT 11
US-10-110-364-16
; Sequence 16, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparenta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 1169505.
US-10-110-364-16

Query Match 100.0%; Score 39; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 EVPGSQH 7
1111111

Db 72 EVPGSQH 78

RESULT 12
US-10-110-364-18
; Sequence 18, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparenta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: US/10/110.364
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 145833.
US-10-110-364-18

Query Match 100.0%; Score 39; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 EVPGSQH 7
1111111
Db 72 EVPGSQH 78

RESULT 13
US-10-110-364-19
; Sequence 19, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparenta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: US/10/110.364
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 1648865
US-10-110-364-19

Query Match 100.0%; Score 39; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.96;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EVPGSOH 7
Db 72 EVPGSOH 78

RESULT 14
US-10-110-364-21

; Sequence 21, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACIBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 408996.
US-10-110-364-21

Query Match 100.0%; Score 39; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSOH 7
Db 72 EVPGSOH 78

RESULT 15
US-10-110-364-23

; Sequence 23, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACIBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 69630.
US-10-110-364-23

Query Match 100.0%; Score 39; DB 6; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EVPGSOH 7
Db 72 EVPGSOH 78

Search completed: July 3, 2002, 09:04:12
Job time: 652 sec

Mon Jul 8 07:51:43 2002

us-09-786-648-2.rapn

Page 6

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 08:52:40 ; Search time 46.57 Seconds
(without alignments)
14.443 Million cell updates/sec

Title: US-09-786-648-2

Perfect score: 39

Sequence: 1 EYPGSGH 7

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	100.0	124	1 XVCB	Cholera enterotoxin
2	39	100.0	124	1 QLECB	heat-labile entero
3	33	84.6	272	2 T14755	hypothetical prote
4	33	84.6	367	2 T01751	gibberellin 20-oxi
5	32	82.1	137	2 G96017	conserved hypotet
6	32	82.1	455	1 A69753	glucuronate dehydrat
7	32	82.1	472	2 AG0432	glutamate synthase
8	32	82.1	534	2 S44886	ZK112.1 protein -
9	31	79.5	137	2 B27586	hypothetical prote
10	31	79.5	242	2 T34767	hypothetical prote
11	31	79.5	260	2 C96827	protein F20B17.2 (
12	31	79.5	273	2 B48820	homeobox protein (
13	31	79.5	274	2 G84353	hypothetical prote
14	31	79.5	353	2 H75446	(S)-2-hydroxy-acid
15	31	79.5	374	2 D81715	conserved hypotet
16	31	79.5	414	2 A37133	apolipoprotein A-I
17	31	79.5	414	2 I48975	hypothetical prote
18	31	79.5	432	2 T46725	cholesterol synthas
19	31	79.5	432	2 T51020	cholesterol synthas
20	31	79.5	459	2 D34791	interleukin-7 rece
21	31	79.5	489	2 T26069	hypothetical prote
22	31	79.5	703	2 AG0242	probable membrane
23	31	79.5	708	2 AG2315	hypothetical prote
24	31	79.5	755	2 D95842	probable beta-gala
25	31	79.5	1028	2 A96719	hypothetical prote
26	31	79.5	1228	2 C98219	proline dehydrogen
27	31	79.5	1228	2 AG3067	hypothetical prote
28	31	79.5	1258	2 T29041	hypothetical prote
29	31	79.5	1259	2 A43425	Bravo/Nr-CAM cell

30	30	76.9	148	2 B72782	hypothetical prote
31	30	76.9	153	2 T31701	hypothetical prote
32	30	76.9	182	2 D83638	conserved hypotet
33	30	76.9	262	2 C34791	interleukin-7 rece
34	30	76.9	270	2 D83072	conserved hypotet
35	30	76.9	291	2 A12241	hypothetical prote
36	30	76.9	298	2 B34791	interleukin-7 rece
37	30	76.9	304	2 F84169	hypothetical prote
38	30	76.9	336	2 A72247	DNA-directed RNA p
39	30	76.9	373	2 T47715	probable 4'-carboxy
40	30	76.9	375	2 T35015	probable 3'-oxadip
41	30	76.9	376	1 S17246	cholesterol synthas
42	30	76.9	412	2 T47142	hypothetical prote
43	30	76.9	431	2 A89761	hypothetical prote
44	30	76.9	436	2 B70321	flavocytchrome C
45	30	76.9	440	2 T44138	hypothetical prote

ALIGNMENTS

RESULT 1
XVCB
Cholera enterotoxin, B chain precursor VC1456 [validated] - Vibrio cholerae (strain N
N: Alternate names: enterotoxin beta chain
C: Species: Vibrio cholerae
C: Date: 24-Apr-1984 #sequence: revision 01-Sep-2000 #text: change 02-Feb-2001
C: Accession: S14624; S39238; S39241; H82196; JCI078; S17666; PC1010; A05130; A01819,
R: Dams, E.; de Wolf, M.; Dierick, W.
submitted to the EMBL Data Library, March 1991
A: Description: Correction of the cholera toxin nucleotide sequence of the Vibrio chol
A: Reference number: S14623
A: Accession: S14624
A: Molecule type: DNA
A: Residues: 1-124 <DM>
A: Cross-references: EMBL:X58786; NID:G48420; PIDN:CAA41593.1; PID:G48422
A: Experimental source: strain 2125
R: Lebens, M.; Holmgren, J.
submitted to the EMBL Data Library, November 1993
A: Description: Structure and arrangement of the Cholera toxin genes in vibrio Cholera
A: Reference number: S39238
A: Accession: S39238
A: Molecule type: DNA
A: Residues: 1-124 <LEB>
A: Cross-references: EMBL:X76390; NID:G433856; PIDN:CAA53973.1; PID:G433857
A: Accession: S39241
A: Molecule type: DNA
A: Residues: 1-124 <LEM>
A: Cross-references: EMBL:X76391; NID:G433859; PIDN:CAA53976.1; PID:G433861
R: Heideberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson,
Chardson, D.; Ermoleva, M.D.; Vamathevan, J.; Baas, S.; Qin, H.; Dragol, I.; Sellc
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A: Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae
A: Reference number: A82035; MID:20406833
A: Accession: H82196
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-124 <HEI>
A: Cross-references: GB:AE004224; GB:AE003852; NID:G9655952; PIDN:AAF94613.1; GSP
R: Shi, C.H.; Cao, C.; Zhang, J.S.; Ma, Q.J.
Chinese Biochem. J. 9, 395-399, 1993
A: Title: Nucleotide sequence analysis of the gene encoding the classical biotype
A: Reference number: JCI078
A: Accession: JCI078
A: Molecule type: DNA
A: Residues: 1-20, 'Q', 22-31, 'Q', 33-38, 'H', 40-49, 'G', 51-67, 'T', 69-124 <SHI>
A: Experimental source: classical biotype strain 569B
R: Dams, E.; de Wolf, M.; Dierick, W.
Biochim. Biophys. Acta 1090, 139-141, 1991
A: Title: Nucleotide sequence analysis of the CT operon of the Vibrio cholerae
A: Reference number: S17665; MID:91355224

A:Accession: S17666
 A:Molecule type: DNA
 A:Residues: 1-38, 'H', 40-67, 'T', 69-124 <DA2>
 A:Cross-references: EMBL:X58785; NID:948888; PIDN:CAA41591.1; PID:g48890
 R:Ma, O.J.; Liu, C.X.; Xiong, L.S.; Yu, X.Q.
 Sci. Sin. B Chem. Biol. Agric. Med. Earth Sci. 34, 274-280, 1991
 A:Title: B subunit of cholera toxin produced in *Escherichia coli*.
 A:Reference number: PC1010
 A:Accession: PC1010
 A:Molecule type: Protein
 A:Residues: 22-38, 'H', 40-41 <MAO>
 R:Meekalanos, J.J.; Swartz, D.J.; Pearson, G.D.N.; Harford, N.; Groyne, F.; de Wilde, M.
 Nature 306, 551-557, 1983
 A:Reference number: A93320; MUID:84068199
 A:Accession: A05130
 A:Molecule type: DNA
 A:Residues: 1-32, 'S', 34-74, 'S', 76-124 <MEK>
 A:Cross-references: GB:X00171; NID:948347; PIDN:CAA24996.1; PID:g758351
 R:Kutrosky, A.; Markel, D.E.; Peterson, J.W.
 J. Biol. Chem. 252, 7257-7264, 1977
 A:Title: Covalent structure of the beta chain of cholera enterotoxin.
 A:Reference number: A01819; MUID:78005537
 A:Accession: A01819
 A:Molecule type: Protein
 A:Residues: 22-38, 'H', 40-42, 'N', 44-67, 'T', 69-90, 'N', 92-124 <KUR>
 R:Lat, C.Y.
 J. Biol. Chem. 252, 7249-7256, 1977
 A:Title: Determination of the primary structure of cholera toxin B subunit.
 A:Reference number: A38033; MUID:78005536
 A:Accession: A38033
 A:Molecule type: Protein
 A:Residues: 22-38, 'H', 40-42, 'N', 44-67, 'T', 69, 'E', 71-90, 'N', 92-124 <LAI>
 A:Note: the difference at residue 70 may be due to deamidation during preparation
 R:Nakashima, Y.; Napiorkowski, P.; Schafer, D.E.; Konigsberg, W.H.
 FEBS Lett. 68, 275-278, 1976
 A:Title: Primary structure of the B subunit of cholera enterotoxin.
 A:Reference number: A38034; MUID:77026365
 A:Accession: A38034
 A:Molecule type: Protein
 A:Residues: 22-38, 'H', 40-67, 'T', 69, 'E', 71, 'QS', 74-75, 'VE', 78-86, 'Q', 88-99, 'Q', 101-103, 'Q'
 R:Takao, T.; Matanabe, H.; Shimonishi, Y.
 Eur. J. Biochem. 146, 503-508, 1985
 A:Title: Facile identification of protein sequences by mass spectrometry.
 A:Reference number: A21910; MUID:85126976
 A:Accession: A21910
 A:Molecule type: Protein
 A:Residues: 22-38, 'H', 40-42, 'N', 44-67, 'T', 69-90, 'N', 92-124 <TAK>
 A:Experimental source: biotype Inaba 569B
 A:Note: Asn-65 was partially deaminated to Asp
 C:Comment: The authors translated the codon TCA for residue 33 as Tyr.
 C:Genetics:
 A:Gene: VC1456
 A:Map position: 1
 C:Complex: the cholera enterotoxin molecule contains three kinds of chains; an alpha and
 ciate noncovalently with the subunit B, an aggregate of five beta chains
 C:Function:
 A:Description: involved in binding of the toxin to cell membranes
 C:Superfamily: cholera enterotoxin beta chain
 C:Keywords: enterotoxin; toxin
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-124/Product: cholera enterotoxin chain B #status predicted <MAT>
 F:30-107/Disulfide bonds: #status experimental

Query Match 100.0%; Score 39; DB 1; Length 124;
 Best Local Similarity 100.0%; Pred. No. 0.5;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
 |||||
 DB 72 EVPGSOH 78

RESULT 2
 OLECB
 heat-labile enterotoxin chain B precursor - *Escherichia coli*
 C:Species: *Escherichia coli*
 C:Date: 29-Jun-1981 #sequence revision 29-Jun-1981 #text change 18-Jun-1999
 C:Accession: A01820; B26946; I41194; I41287; I67644; A61475
 R:Dallas, W.S.; Falkow, S.
 Nature 288, 499-501, 1980
 A:Title: Amino acid sequence homology between cholera toxin and *Escherichia coli*
 A:Reference number: A01820; MUID:81074965
 A:Accession: A01820
 A:Molecule type: mRNA
 A:Residues: 1-124 <DAL>
 R:Yamamoto, T.; Gojobori, T.; Yokota, T.
 J. Bacteriol. 169, 1352-1357, 1987
 A:Title: Evolutionary origin of pathogenic determinants in enterotoxigenic *Escherichia coli*
 A:Reference number: A26946; MUID:87137303
 A:Accession: B26946
 A:Molecule type: DNA
 A:Residues: 1-27, 'E', 29-63, 'K', 65-124 <YAN>
 A:Cross-references: EMBL:M15363; NID:g148335; PIDN:AAA24792.1; PID:g148336
 R:Leong, J.; Vinal, A.C.; Dallas, W.S.
 Infect. Immun. 48, 73-77, 1985
 A:Title: Nucleotide sequence comparison between heat-labile toxin B-subunit cistrons
 A:Reference number: I41194; MUID:85156481
 A:Accession: I41194
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-5, 'F', 7-17, 'C', 19-24, 'S', 26-27, 'E', 29-33, 'H', 35-63, 'K', 65-66, 'A', 68-122
 A:Cross-references: GB:M17874; NID:g145830; PIDN:AAA98064.1; PID:g145831
 R:Brabham, I.; Gentz, R.
 J. Biol. Chem. 262, 10189-10194, 1987
 A:Title: A functional interaction between the signal peptide and the translation apparatus
 A:Reference number: I41287; MUID:87280041
 A:Accession: I41287
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-22 <RE2>
 A:Cross-references: GB:M17101; NID:g146375; PIDN:AAA23973.1; PID:g146376
 R:Imoue, T.; Tsuji, T.; Koto, M.; Imamura, S.; Miyama, A.
 FEBS Microbiol. Lett. 108, 157-161, 1993
 A:Title: Amino acid sequence of heat-labile enterotoxin from chicken enterotoxigenic
 A:Reference number: I5342; MUID:93252225
 A:Accession: I67644
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-17, 'C', 19, 'Y', 21-24, 'S', 26-27, 'E', 29-63, 'K', 65-66, 'A', 68-122, 'E', 124 <R
 A:Cross-references: GB:S60731; NID:9408994; PIDN:AA60441.1; PID:9408996
 R:Tsuji, T.; Lida, T.; Honda, T.; Mawatani, T.; Nagahama, M.; Sakurai, J.; Wada, K.
 Microb. Pathog. 2, 381-390, 1987
 A:Title: A unique amino acid sequence of the B subunit of a heat-labile enterotoxin 1
 A:Reference number: A61475; MUID:89180953
 A:Accession: A61475
 A:Molecule type: Protein
 A:Residues: 22-24, 'S', 26-27, 'E', 29-63, 'K', 65-66, 'A', 68-95, 'A', 97-122, 'E', 124 <TSU>
 A:Experimental source: strain 240-3
 C:Complex: the heat-labile enterotoxin molecule contains one A chain and five or six
 C:Function:
 A:Description: the biological activity of the toxin is produced by the A chain, which
 C:Superfamily: cholera enterotoxin beta chain
 C:Keywords: enterotoxin
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-124/Product: heat-labile enterotoxin chain B #status predicted <MAT>
 F:30-107/Disulfide bonds: #status predicted

Query Match 100.0%; Score 39; DB 1; Length 124;
 Best Local Similarity 100.0%; Pred. No. 0.5;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7

Db 72 EVPGSQH 78

RESULT 3

T14755

hypothetical protein DKFZP564A0122.1 - human

C:Species: Homo sapiens (man)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T14755

R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, August 1999

A:Reference number: Z18181

A:Accession: T14755

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-272 <NMB>

A:Cross-references: EMBL:AL110209

A:Experimental source: fetal brain; clone DKFZP564A0122

A:Note: DKFZP564A0122.1

Query Match

Best Local Similarity 84.6%; Score 33; DB 2; Length 272;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSQH 7

Db 246 EVPGSEH 252

RESULT 4

T01751

giberellin 20-oxidase - common tobacco

N:Alternate names: Ntc16 protein

C:Species: Nicotiana tabacum (common tobacco)

C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 20-Jun-2000

C:Accession: T01751

R:Tanaka-Ieguchi, M.; Itoh, H.; Oyama, N.; Koshioke, M.; Matsuo, M.
submitted to the EMBL Data Library, July 1998

A:Description: Over-expression of a tobacco homeobox gene, NTH15, decreases the expression

A:Reference number: Z14418

A:Accession: T01751

A:Status: preliminary; translated from GB/EMBL/DDB

A:Molecule type: mRNA

A:Residues: 1-367 <TAN>

A:Cross-references: EMBL:AB016084

C:Genetics:

A:Gene: NTC16

C:Superfamily: 1-aminocyclopropane-1-carboxylate oxidase

Query Match

Best Local Similarity 84.6%; Score 33; DB 2; Length 367;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSQH 7

Db 149 EVPGSQH 155

RESULT 5

G96017

conserved hypothetical exported protein SMD20700 [imported] - Sinorhizobium meliloti (str)

C:Species: Sinorhizobium meliloti

C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C:Accession: G96017

R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Hernan

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1,683-kb PSYMB megaplasmid from the N2-fixing endo

A:Reference number: A95842; MUID:21396508; PMID:11481431

A:Accession: G96017

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-137 <KUR>

A:Cross-references: GB:AL591985; PIDN:CAC49807.1; PID:q15141295; GSPDB:GN00167

A:Experimental source: strain 1021, megaplasmid PSYMB

R:Galibert, F.; Finan, T.M.; Long, S.R.; Pulner, A.; Abola, P.; Ampe, F.; Barloy-Hubli

pell, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau

hebut, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,

A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: SMD20700

A:Genome: plasmid

Query Match

Best Local Similarity 82.1%; Score 32; DB 2; Length 137;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSQH 7

Db 98 EVPGSQH 104

RESULT 6

A69753

glucarate dehydratase (EC 4.2.1.40) - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000

C:Accession: A69753

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fuma, S.; Galliz, A.; Gal

lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M

Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma

Y, M.; Ogawa, K.; Ogata, A.; Oudega, B.; Park, S.H.; Paro, V.; Pohl, T.M.; Portee

Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl

A:Authors: Schleich, S.; Schroeter, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Se

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis

A:Reference number: A69580; MUID:98044033

A:Accession: A69753

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-455 <KUN>

A:Cross-references: GB:299105; GB:AL009126; NID:q2632457; PIDN:CAB12043.1; PID:q26325

A:Experimental source: strain 168

C:Genetics:

A:Gene: ycbF

C:Superfamily: glucarate dehydratase

C:Keywords: carbon-oxygen lyase; hydro-lyase

OY 1 EVPGSQH 7

Db 64 EVPGSEH 70

RESULT 7

AG0432

glutamate synthase (NADPH) (EC 1.4.1.13) small chain [imported] - Yersinia pestis (strain C:Species: Yersinia pestis
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AC0432
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarragge, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett, Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: AC0432
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-472 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CAC92787.1; PID:q15981480; GSPDB:GN00175
 C:Genetics:
 A:Gene: gld
 C:Superfamily: glutamate synthase small chain
 C:Keywords: oxidoreductase

Query Match 82.1% Score 32; DB 2; Length 472;
 Best Local Similarity 57.1% Pred. No. 65;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7
 ::|||:|
 Db 382 QIPGSEH 388

RESULT 8
 S4486
 ZK112.1 protein - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 14-Sep-1994 #sequence_revision 12-May-1995 #text_change 23-Mar-2001
 C:Accession: S4486
 R:Du, Z.
 submitted to the EMBL Data Library, May 1993
 A:Description: Sequence of the C. elegans cosmid ZK112.
 A:Reference number: S4486
 A:Accession: S4486
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-534 <DUZ>
 A:Cross-references: EMBL:L14324; NID:q289740; PID:q289741
 C:Genetics:
 A:Introns: 25/3; 65/2; 196/2; 249/1; 275/1; 385/2; 415/2
 C:Superfamily: Caenorhabditis elegans ZK688.6 protein

Query Match 82.1% Score 32; DB 2; Length 534;
 Best Local Similarity 71.4% Pred. No. 74;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYVGSQH 7
 ||:||||
 Db 442 EIPGSAH 448

RESULT 9
 B27586
 hypothetical protein - Mycobacterium leprae
 C:Species: Mycobacterium leprae
 C:Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 26-May-2000
 C:Accession: B27586
 R:Booth, R.J.; Harris, D.P.; Love, J.M.; Watson, J.D.
 J. Immunol. 140, 597-601, 1988
 A:Title: Antigenic proteins of Mycobacterium leprae. Complete sequence of the gene for
 A:Reference number: A92821; MUID:88088878
 A:Accession: B27586
 A:Molecule type: DNA
 A:Residues: 1-137 <BOO>
 A:Cross-references: GB:M19058; NID:q149919; PIDN:AAA88230.1; PID:q1196505

C:Superfamily: Mycobacterium leprae hypothetical 15.2K protein

Query Match 79.5% Score 31; DB 2; Length 137;
 Best Local Similarity 83.3% Pred. No. 28;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 VPGSQH 7
 |||:|
 Db 127 VPGNOH 132

RESULT 10
 T34767
 hypothetical protein SC2A11.21c SC2A11.21c - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
 C:Accession: T34767
 R:Murphy, L.; Harris, D.; Parkhill, J.; Barrett, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, August 1998
 A:Reference number: T34767
 A:Accession: T34767
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-242 <MU>
 A:Cross-references: EMBL:AL031184; PIDN:CAA20190.1; GSPDB:GN00070; SCOEDB:SC2A11.21c
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SC2A11.21c

Query Match 79.5% Score 31; DB 2; Length 242;
 Best Local Similarity 71.4% Pred. No. 52;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYVGSQH 7
 ||||:|
 Db 205 EYVGDH 211

RESULT 11
 C96827
 protein F20B17.2 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: C96827
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
 ansen, N.F.; Hughes, B.; Huizart, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Malt, R.; Marzla
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A66141; MUID:21016719
 A:Accession: C96827
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-260 <STO>
 A:Cross-references: GB:AE005173; NID:g7715588; PIDN:AAF68106.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F20B17.2
 A:Map position: 1

Query Match 79.5% Score 31; DB 2; Length 260;
 Best Local Similarity 71.4% Pred. No. 56;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYVGSQH 7
 :|||:|

DB 66 KVPKOH 72

RESULT 12

B48820

homeobox protein (clone NvHox-4) - eastern newt

C:Species: *Notophthalmus viridescens*, *Triturus viridescens* (eastern newt)

C:date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 24-Sep-1999

C:Accession: B48820

R:Beauchemin, M.; Savard, P.

Dev. Biol. 154, 55-65, 1992

A:Title: Two digital-less related homeobox-containing genes expressed in regeneration bio

A:Reference number: A48820; MUID:93050784

A:Accession: B48820

A:Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-273 <BEA>

A:Cross-references: GB:X63531; GB:S47223; NID:9432377; PIDN:CAA45094.1; PID:9432378

A:Note: sequence extracted from NCBI backbone (NCBIN:117052, NCBIP:117053)

C:Superfamily: unassigned homeobox proteins; homeobox homology

C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

F:126-182/Domain: homeobox homology <HOX>

Query Match

Best Local Similarity 79.5%; Score 31; DB 2; Length 273;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSOH 7

DB 188 EVPGMEH 194

RESULT 13

G84353

hypothetical protein Vng2034h [imported] - *Halobacterium* sp. NRC-1C:Species: *Halobacterium* sp. NRC-1

C:date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: G84353

R:Ng, W.V.; Kennedy, S.P.; Mahatras, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.

; Leihhauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabc

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li

A:Title: Genome sequence of *Halobacterium* species NRC-1.

A:Reference number: A84160; MUID:20504483

A:Accession: G84353

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-274 <STO>

A:Cross-references: GB:AE004437; NID:910581460; PIDN:AMG20195.1; GSPDB:GMO0138

C:Genetics:

A:Gene: VNG2034H

Query Match

Best Local Similarity 79.5%; Score 31; DB 2; Length 274;

Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVPGSOH 7

DB 117 EAPGDH 123

RESULT 14

H75446

(S)-2-hydroxy-acid oxidase - *Deinococcus radiodurans* (strain R1)C:Species: *Deinococcus radiodurans*

C:date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000

C:Accession: H75446

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

M.; Shen, M.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.

A:Reference number: A75250; MUID:20036896

A:Accession: H75446

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-353 <WHT>

A:Cross-references: GB:AE001954; GB:AE000513; NID:96458751; PIDN:AAF10604.1; PID:9645

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR1031

A:Map position: 1

C:Superfamily: (S)-2-hydroxy-acid oxidase; (S)-2-hydroxy-acid oxidase homology

F:3-297/Domain: (S)-2-hydroxy-acid oxidase homology <2HT>

Query Match

Best Local Similarity 79.5%; Score 31; DB 2; Length 353;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 VPQSOH 7

DB 189 VPQSEH 194

RESULT 15

D81715

conserved hypothetical protein TC0328 [imported] - *Chlamydia muridarum* (strain Nigg)C:Species: *Chlamydia muridarum*, *Chlamydia trachomatis* Mohn

C:date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000

C:Accession: D81715

R:Read, T.D.; Brinham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke

, C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe

Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of *Chlamydia trachomatis* Mohn and *Chlamydia pneumoniae* AR39

A:Reference number: A81500; MUID:20150255

A:Accession: D81715

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-374 <TEW>

A:Cross-references: GB:AE002300; GB:AE002160; NID:97190365; PIDN:AAF39192.1; PID:9719

C:Genetics:

A:Gene: TC0328

Query Match

Best Local Similarity 79.5%; Score 31; DB 2; Length 374;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 VPQSOH 7

DB 327 LPQSOH 332

Search completed: July 3, 2002, 08:56:32
Job time: 232 sec

Mon Jul 8 07:51:44 2002

us-09-786-648-2.rpt

Page 6

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 08:55:00 ; Search time 21.51 Seconds
(Without alignments)
12.601 Million cell updates/sec

Title: US-09-786-648-2
Perfect score: 39
Sequence: 1 EVPGSQH 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	100.0	124	1	CHTB_VIRCH
2	39	100.0	124	1	ELBH_ECOLI
3	39	100.0	124	1	ELBP_ECOLI
4	32	82.1	455	1	GUDH_BACSU
5	32	82.1	534	1	YOG1_CAEEL
6	32	82.1	666	1	PD14_MOUSE
7	32	82.1	765	1	SIM1_MOUSE
8	31	79.5	137	1	YH96_MYCLE
9	31	79.5	219	1	CITDA_HUMAN
10	31	79.5	240	1	YB5A_THEMA
11	31	79.5	273	1	DLX3_NORVI
12	31	79.5	274	1	DLX3_PLEMA
13	31	79.5	280	1	DLX3_AMBME
14	31	79.5	414	1	COT2_BOVIN
15	31	79.5	414	1	COT2_HUMAN
16	31	79.5	414	1	COT2_MOUSE
17	31	79.5	414	1	COT2_RAT
18	31	79.5	432	1	AROC_NEUCR
19	31	79.5	459	1	IL7R_MOUSE
20	31	79.5	710	1	IRAL_MOUSE
21	31	79.5	743	1	BGAL_THRET
22	30	76.9	182	1	KPTA_PSEAE
23	30	76.9	336	1	RPOA_THEMA
24	30	76.9	376	1	AROC_YEAST
25	30	76.9	459	1	IL7R_HUMAN
26	30	76.9	500	1	GABT_HUMAN
27	30	76.9	504	1	A37C_DROME
28	30	76.9	573	1	AMH2_HUMAN
29	30	76.9	616	1	REFX5_HUMAN
30	30	76.9	622	1	SR68_CAEEL
31	30	76.9	766	1	SIM1_HUMAN
32	30	76.9	775	1	TH1L_SCHPO
33	30	76.9	805	1	AHR_MOUSE

34	30	76.9	808	1	PLD_TOBAC	P93400 nicotiana t
35	30	76.9	853	1	AHR_RAT	P41738 rattus norv
36	30	76.9	853	1	ST14_HUMAN	Q95576 homo sapien
37	30	76.9	954	1	DRP2_HUMAN	Q13474 homo sapien
38	30	76.9	1286	1	PATC_DROME	P18502 drosophila
39	30	76.9	1538	1	LHR_ECOLI	P30015 escherichia
40	30	76.9	1638	1	BRM_DROME	P25439 drosophila
41	30	76.9	1808	1	TENA_CHICK	P10039 gallus gall
42	29	74.4	277	1	ZDKG_CONSP	P15339 corynebacte
43	29	74.4	395	1	INX3_DROME	Q9vas7 drosophila
44	29	74.4	404	1	Y4XM_RHISN	P55705 rhizobium s
45	29	74.4	586	1	RRPO_BWYVF	P09507 beet wester

ALIGNMENTS

RESULT 1	ID	CHTB_VIRCH	STANDARD:	PRT:	124 AA.
AC	P01556	Q9JQ02			
DT	21-JUL-1986	(Rel. 01, Created)			
DT	13-AUG-1987	(Rel. 05, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Cholera enterotoxin, beta chain precursor.				
GN	CITX OR TOXB OR VC1456.				
OS	Vibrio cholerae.				
OC	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.				
OX	NCBI_Taxid=666;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=84061784; PubMed=6315707;				
RA	Lockman H., Kaper J.B.;				
RT	"Nucleotide sequence analysis of the A2 and B subunits of Vibrio				
RL	cholerae enterotoxin.";				
RL	J. Biol. Chem. 258:13722-13726(1983).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	STRAIN=EL TOR 2125;				
RC	MEDLINE=84068199; PubMed=6646234;				
RA	Mekalanos J.J., Swartz D.J., Pearson G.D.N., Harford N., Groyne F.,				
RT	de Wilde M.;				
RT	"Cholera toxin genes: nucleotide sequence, deletion analysis and				
RT	vaccine development.";				
RL	Nature 306:551-557(1983).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=EL TOR 2125;				
RA	Dans E., de Wolf M., Dierick W.;				
RL	Submitted (May-1991) to the EMBL/GenBank/DBJ databases.				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RX	STRAIN=4260B / SEROTYPE O139;				
RC	MEDLINE=94237453; PubMed=8181723;				
RA	Lebens M., Holmgren J.;				
RT	"Structure and arrangement of the cholera toxin genes in Vibrio				
RT	cholerae O139.";				
RL	FEMS Microbiol. Lett. 117:197-202(1994).				
RN	[5]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=1854 / O139-BENGAL;				
RA	Yamamoto K., Do V.G.R.F., Xu M., Iida T., Miwatani T., Albert M.J.,				
RL	Honda T.;				
RL	Submitted (May-1994) to the EMBL/GenBank/DBJ databases.				
RN	[6]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=EL TOR N16961 / SEROTYPE O1;				
RX	MEDLINE=20406833; PubMed=10952301;				
RA	Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,				
RA	Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,				
RA	Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,				
RA	Ernolleva M.D., Vamathevan J., Bass S., Qin H., Dragoti I., Sellers P.,				
RA	McDonald L., Uitterback T., Fleischmann R.D., Nierman W.C., White O.,				

RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 RT cholerae";
 RL Nature 406:477-483(2000).
 RN [7]
 RP SEQUENCE OF 22-124.
 RX MEDLINE=78005537; PubMed=903363;
 RA Kurosky A., Markel D.E., Peterson J.W.;
 RT "Covalent structure of the beta chain of cholera enterotoxin.";
 RL J. Biol. Chem. 252:7257-7264(1977).
 RN [8]
 RP SEQUENCE OF 22-124.
 RX MEDLINE=78005536; PubMed=903362;
 RA Lai C.-Y.;
 RT "Determination of the primary structure of cholera toxin B subunit.";
 RL J. Biol. Chem. 252:7249-7256(1977).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RX MEDLINE=94272319; PubMed=8003934;
 RA Merritt E.A., Sarfaty S., van den Akker F., L'Hoir C., Martini J.A.,
 RT "Crystal structure of cholera toxin B-pentamer bound to receptor GM1
 RT pentasaccharide.";
 RL Protein Sci. 3:166-175(1994).
 RN [10]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
 RX MEDLINE=95387394; PubMed=7658472;
 RA Zhang R.G., Westbrook M.L., Westbrook E.M., Scott D.L., Orlowski Z.,
 RT "The 2.4 A crystal structure of cholera toxin B subunit pentamer:
 RT choleragenoid.";
 RL J. Mol. Biol. 251:550-562(1995).
 RN [11]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RC STRAIN-OGAWA 41 / CLASSICAL BIOTYPE;
 RX MEDLINE=97376625; PubMed=9232653;
 RA Merritt E.A., Sarfaty S., Jobling M.G., Chang T., Holmes R.K.,
 RT "Structural studies of receptor binding by cholera toxin mutants.";
 RL Protein Sci. 6:1516-1528(1997).
 CC -I- FUNCTION: THE BETA CHAIN AGGREGATE (B SUBUNIT) IS INVOLVED IN
 CC BINDING TO CELL MEMBRANES.
 CC -I- SUBUNIT: CONTAINS 3 KINDS OF CHAINS. AN ALPHA AND A GAMMA CHAIN
 CC (FROM THE SAME PRECURSOR MOLECULE), LINKED BY AN INTERCHAIN
 CC DISULFIDE BOND, ASSOCIATE NONCOVALENTLY WITH AN AGGREGATE OF 4 TO
 CC 6 BETA CHAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X00171; CAA24996.1; -
 DR EMBL: K01170; AAA27573.1; -
 DR EMBL: D30053; BAA06291.1; -
 DR EMBL: X58786; CAA41593.1; -
 DR EMBL: X76390; CAA53973.1; -
 DR EMBL: X76391; CAA53976.1; -
 DR EMBL: AE004224; AAF94613.1; -
 DR PIR: A01819; XVCB.
 DR PIR: A05130; A05130.
 DR PIR: S14624; S14624.
 DR PDB: 2CHB; 03-DEC-97.
 DR PDB: 3CHB; 12-AUG-98.
 DR PDB: 1CHP; 08-MAR-96.
 DR PDB: 1CHO; 08-MAR-96.
 DR PDB: 1FGB; 23-DEC-96.
 DR PDB: 1XPB; 01-APR-97.
 DR PDB: 1XTC; 01-AUG-96.

DR PDB: 1CT1; 15-OCT-97.
 DR TIGR: VCI456; -
 DR InterPro: IPR001835; Enterotoxin_B.
 DR Pfam: PF01376; Enterotoxin_B; 1.
 DR PRINTS: PR00772; ENTEROTOXINB.
 DR ProDom: PD012805; Enterotoxin_B; 1
 KW Membrane; Enterotoxin; Signal; 3D-structure; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 124
 FT DISULFID 30 107
 FT CONFLICT 33 33
 FT CONFLICT 39 39
 FT CONFLICT 43 43
 FT CONFLICT 68 68
 FT CONFLICT 70 70
 FT CONFLICT 75 75
 FT CONFLICT 91 91
 FT HELIX 26 30
 FT TURN 31 32
 FT TURN 34 35
 FT STRAND 36 44
 FT STRAND 47 51
 FT STRAND 54 55
 FT STRAND 58 62
 FT TURN 64 65
 FT STRAND 68 71
 FT TURN 76 77
 FT TURN 80 99
 FT TURN 100 100
 FT STRAND 102 109
 FT STRAND 115 123
 SQ SEQUENCE 124 AA; 13957 MW; 9AA393E3EA8E3EBF CRC64;
 Query Match 100.0%; Score 39; DB 1; Length 124;
 Best Local Similarity 100.0%; Pred No. 0.23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EYVPSQH 7
 Db 72 EYVPSQH 78

RESULT 2
 ELBH.ECOLI
 ID ELBH.ECOLI STANDARD; PRT; 124 AA.
 AC P13811;
 DT 01-JAN-1990 (rel. 13, Created)
 DT 01-JAN-1990 (rel. 13, Last sequence update)
 DT 15-JUL-1999 (rel. 38, Last annotation update)
 DE Heat-labile enterotoxin B chain precursor (LT-B, human) (LTH-B).
 GN ELTB OR LTBP.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 OX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ISOLATE H74-114;
 RX MEDLINE=85156481; PubMed=3884513;
 RA Leong J., Vinal A.C., Dallas W.S.;
 RT "Nucleotide sequence comparison between heat-labile toxin B-subunit
 RT cistrons from Escherichia coli of human and porcine origin.";
 RL Infect. Immun. 48:73-77(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ISOLATE H10407;
 RX MEDLINE=83114628; PubMed=6759877;
 RA Yamamoto T., Tamura T.A., Yokota T., Takano T.;
 RT "Overlapping genes in the heat-labile enterotoxin operon originating
 RT from Escherichia coli human strain.";
 RL Mol. Gen. Genet. 188:356-359(1982).
 RN [3]

RP SEQUENCE FROM N.A.
 RC STRAIN-ISOLATE H10407;
 RX MEDLINE=93252225; PubMed=8486242;
 RA Inoue T., Tsuji T., Koto M., Imamura S., Miyama A.;
 RT "Amino acid sequence of heat-labile enterotoxin from chicken
 RT enterotoxigenic Escherichia coli is identical to that of human strain
 RT H 10407.";
 RL FEMS Microbiol. Lett. 108:157-161(1993).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ETEC LT 87;
 RA Germani Y., Desperlier J.M.;
 RL Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
 RN [5]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE=95349400; PubMed=7623669;
 RA Domenighini M., Pizza M., Jobling M.G., Holmes R.K., Rappuoli R.;
 RT "Identification of errors among database sequence entries and
 RT comparison of correct amino acid sequences for the heat-labile
 RT enterotoxins of Escherichia coli and Vibrio cholerae.";
 RL Mol. Microbiol. 15:1165-1167(1995).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (3.04 ANGSTROMS).
 RX MEDLINE=99185101; PubMed=10085117;
 RA Matkovic-Galogovic D., Loregian A., D'Acunzio M.R., Battistutta R.,
 RA Tossi A., Palu G., Zanotti G.;
 RT "Crystal structure of the B subunit of Escherichia coli heat-labile
 RT enterotoxin carrying peptides with anti-herpes simplex virus type 1
 RT activity.";
 RL J. Biol. Chem. 274:8764-8769(1999).
 CC -1- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
 CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
 CC -1- SUBUNIT: HETEROHXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M17874; AAA98064.1; -;
 DR EMBL: J01646; AAB02982.1; -;
 DR EMBL: S60731; AAC60441.1; -;
 DR EMBL: X83966; CAA58800.1; -;
 DR PDB: 1LTR; 23-MAR-99.
 DR InterPro: IPR001835; Enterotoxin_B.
 DR Pfam: PF01376; Enterotoxin_B; 1.
 DR PRINTS: PR00772; ENTEROTOXINB.
 DR ProDom: PD012805; Enterotoxin_B; 1.
 DR Enterotoxin; Signal; 3D-structure.
 FT SIGNAL 1 21
 FT CHAIN 22 124 HEAT-LABILE ENTEROTOXIN B CHAIN.
 FT DISULFID 30 107
 FT VARIANT 6 6 F -> C (IN ISOLATE H10407).
 FT VARIANT 20 20 H -> Y (IN ISOLATE H10407).
 FT VARIANT 34 34 H -> R (IN ISOLATE H10407).
 SO SEQUENCE 124 AA; 14027 MW; E9F7FC7B9D3BC47 CRC64;

Query Match 100.0%; Score 39; DB 1; Length 124;
 Best Local Similarity 100.0%; Pred. No. 0.23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7
 DB 72 EYVGSQH 78

RESULT 3
 EIBP_ECOLI STANDARD; PRT: 124 AA.

AC P32890; P13768; P01557;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 GN Heat-labile enterotoxin B chain precursor (LT-B, porcine) (LTP-B).
 GN EIBP OR LTPB.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxId=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ISOLATE P307;
 RX MEDLINE=81074965; PubMed=7003397;
 RA Dallas W.S., Falkow S.;
 RT "Amino acid sequence homology between cholera toxin and Escherichia
 RT coli heat-labile toxin.";
 RL Nature 288:499-501(1980).
 RN [2]
 RP REVISIONS TO 28 AND 64.
 RC STRAIN-ISOLATE P307;
 RX MEDLINE=85156481; PubMed=3884513;
 RA Leong J., Vinal A.C., Dallas W.S.;
 RT "Nucleotide sequence comparison between heat-labile toxin B-subunit
 RT cistrons from Escherichia coli of human and porcine origin.";
 RL Infect. Immun. 48:73-77(1985).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ISOLATE PCG86;
 RX MEDLINE=87137303; PubMed=3546273;
 RA Yamamoto T., Gojohori T., Yokota T.;
 RT "Evolutionary origin of pathogenic determinants in enterotoxigenic
 RT Escherichia coli and Vibrio cholerae O1.";
 RL J. Bacteriol. 169:1352-1357(1987).
 RN [4]
 RP SEQUENCE OF 1-22 FROM N.A.
 RX MEDLINE=87280041; PubMed=3301830;
 RA Ibrahim I., Gentr R.;
 RT "A functional interaction between the signal peptide and the
 RT translation apparatus is detected by the use of a single point
 RT mutation which blocks translocation across mammalian endoplasmic
 RT reticulum.";
 RL J. Biol. Chem. 262:10189-10194(1987).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).
 RX MEDLINE=93240541; PubMed=8478941;
 RA Sixma T.K., van Zanten B.A.M., Dauter Z., Hol W.G.J.;
 RT "Refined structure of Escherichia coli heat-labile enterotoxin, a
 RT close relative of cholera toxin.";
 RL J. Mol. Biol. 230:890-918(1993).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
 RX MEDLINE=91238966; PubMed=2034287;
 RA Sixma T.K., Pronk S.E., Kalk K.H., Wartna E.S., van Zanten B.A.M.,
 RA Witolt B., Hol W.G.J.;
 RT "Crystal structure of a cholera toxin-related heat-labile enterotoxin
 RT from E. coli.";
 RL Nature 351:371-377(1991).
 RN [7]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE=95349400; PubMed=7623669;
 RA Domenighini M., Pizza M., Jobling M.G., Holmes R.K., Rappuoli R.;
 RT "Identification of errors among database sequence entries and
 RT comparison of correct amino acid sequences for the heat-labile
 RT enterotoxins of Escherichia coli and Vibrio cholerae.";
 RL Mol. Microbiol. 15:1165-1167(1995).
 CC -1- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
 CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
 CC -1- SUBUNIT: HETEROHXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M17873; AAA98065.1; -;
DR EMBL: M15363; AAA24792.1; -;
DR EMBL: M17101; AAA23973.1; -;
DR PIR: A01820; QLECB.
DR PIR: B26946; QLECB.
DR PDB: 1LTA; 31-JAN-94.
DR PDB: 1LTA; 31-JAN-94.
DR PDB: 1LTA; 15-SEP-95.
DR PDB: 1LTA; 17-AUG-96.
DR PDB: 1LTA; 31-JAN-94.
DR PDB: 1LTA; 31-JAN-94.
DR PDB: 1LTA; 07-JUL-97.
DR PDB: 1LTA; 16-JUN-97.
DR PDB: 1LTA; 03-DEC-97.
DR PDB: 1LTA; 03-DEC-97.
DR PDB: 1LTA; 20-APR-95.
DR InterPro: IPR01835; Enterotoxin_B.
DR Pfam: PF01376; Enterotoxin_B. 1.
DR ProDom: PD012805; Enterotoxin_B. 1.
DR Enterotoxin; Signal; 3D-structure.
KW SIGNAL 1 21
FT CHAIN 22 124 HEAT-LABILE ENTEROTOXIN B CHAIN.
FT DISULFID 30 107
FT HELIX 26 30
FT TURN 31 32
FT STRAND 36 43
FT STRAND 47 51
FT TURN 54 55
FT STRAND 58 62
FT TURN 64 65
FT STRAND 68 71
FT TURN 76 77
FT HELIX 80 98
FT TURN 99 100
FT STRAND 103 109
FT STRAND 115 123
SQ SEQUENCE 124 AA; 14133 MW; 6DB7DE58395EA70D CRC64;

Query Match 100.0%; Score 39; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSQH 7
Db 72 EVPGSQH 78

RESULT 4
GUDH_BACSU STANDARD: PRT; 455 AA.
ID GUDH_BACSU
AC P42238;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable glucarate dehydratase (EC 4.2.1.40) (GUDH) (glucD).
GN GUDH.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
SEQUENCE FROM N.A. *
RP STRAIN=168;
RX MEDLINE=95219079; PubMed=7704254;
RA Ogawa K.-I., Akagawa E., Nakamura K., Yamane K.;
RT "Determination of a 21548 bp nucleotide sequence around the 24

RT degrees region of the Bacillus subtilis chromosome.";
RL Microbiology 141:269-275(1995).

CC -1- FUNCTION: CATALYZES THE DEHYDRATION OF GLUCARATE TO 5-KETO-4-
CC DEOXY-D-GLUCARATE (5-KDGLUC) (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: D-glucarate + 5-dehydro-4-deoxy-D-glucarate +
CC H₂O.
CC -1- PATHWAY: FIRST STEP IN GLUCARATE CATABOLISM.
CC -1- SIMILARITY: BELONGS TO THE MANDATELASE RACEMASE / MUCONATE
CC LACTONIZING ENZYME FAMILY. GLUCD SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: D30808; BAA06470.1; -;
DR EMBL: Z99105; CAB12043.1; -;
DR HSSP: P42206; 1BQG.
DR Subtilist; BG11161; gudD.
DR InterPro: IPR001354; MR_MLE.
DR Pfam: PF01188; MR_MLE. 1.
DR Pfam: PF02746; MR_MLE_N; 1.
DR Lyase; Complete proteome.
KW SEQUENCE 455 AA; 50782 MW; 3238486007698C2A CRC64;

Query Match 82.1%; Score 32; DB 1; Length 455;
Best Local Similarity 71.4%; Pred. No. 27;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSQH 7
Db 64 EVPGSQH 70

RESULT 5
YOG1_CAEEL STANDARD: PRT; 534 AA.
ID YOG1_CAEEL
AC P34610;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Putative serine protease ZK112.1 precursor (EC 3.4.-.-).
GN ZK112.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copestake T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkhen R.,
RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wooldman P.;
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S28 (SERINE PROTEASE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: L14324; AAA28181.1; -
DR PIR: S44886; S44886.
DR WormPep: ZK112.1; CE00372.
DR InterPro: IPR000073; Abhydrolase.
DR InterPro: IPR000379; Est_Lip_thioest_actsite.
DR Pfam: PF00561; abhydrolase; 1.
DR Hypothetical protein: Hydrolase; Serine protease; Signal.
FT SIGNAL 1 17
FT CHAIN 18 334
FT ACT_SITE 177 177 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 420 420 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 448 448 CHARGE RELAY SYSTEM (POTENTIAL).
FT CARBOHYD 69 69 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 107 107 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 240 240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 244 244 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 534 AA; 59221 MW; C43F0104B42E4630 CRC64;

Query Match 82.1%; Score 32; DB 1; Length 534;
Best Local Similarity 71.4%; Pred. No. 32;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYFGSOH 7
I:| | | |
Db 442 EIRGSAH 448

RESULT 6
PDI4_MOUSE STANDARD; PRT; 666 AA.
AC 092183;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE Protein-arginine deiminase type IV (EC 3.5.3.15) (peptidylarginine deiminase IV).
GN PDI4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Epidermis;
RX MEDLINE=99192810; PubMed=10092850;
RA Rusd A.A., Ikejiri Y., Ono H., Yonekawa T., Shiraiwa M., Kawada A., Takahara H.;
RT "Molecular cloning of cDNAs of mouse peptidylarginine deiminase type I, type III and type IV, and the expression pattern of type I in mouse."
RT Eur. J. Biochem. 259:660-669(1999).
RT -I- FUNCTION: CATALYZES THE DEIMINATION OF ARGININE RESIDUES OF PROTEINS.
RT -I- CATALYTIC ACTIVITY: Protein L-arginine + H(2)O = protein L-citrulline + NH(3).
RT -I- COFACTOR: REQUIRES CALCIUM IONS.
RT -I- SIMILARITY: SOME SIMILARITY WITH ARGINASE AND TO ARGININE DEIMINASE.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: AB013850; BAA34246.1; -
DR MGD: MGI:1338898; PDI4.
DR InterPro: IPR004303; PAD.
DR Pfam: PF03068; PAD; 1.
DR Hydrolase; Calcium-binding; Multigene family.
FT CA_BIND 505 516 EF-HAND (POTENTIAL).
SQ SEQUENCE 666 AA; 74476 MW; 70FAE4E7E232D34A CRC64;

Query Match 82.1%; Score 32; DB 1; Length 666;
Best Local Similarity 71.4%; Pred. No. 40;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYFGSOH 7
I:| | | |
Db 239 ELPGSOH 245

RESULT 7
SIM1_MOUSE STANDARD; PRT; 765 AA.
AC 061045; P70183;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Single-minded homolog 1 (SIM1).
GN SIM1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS WEBSTER;
RX MEDLINE=97020303; PubMed=8812055;
RA Fan C.-M., Kuwana E., Bulfone A., Fletcher C.F., Copeland N.G., Jenkins N.A., Crews S., Martinez S., Puellès L., Rubenstein J.L., Tessier-Lavigne M.;
RT "Expression patterns of two murine homologs of Drosophila single-minded suggest possible roles in embryonic patterning and in the pathogenesis of Down syndrome."
RT Mol. Cell. Neurosci. 7:1-16(1996).
RN [2]
RP ERRATUM.
RX MEDLINE=97029422; PubMed=8875433;
RA Fan C.-M., Kuwana E., Bulfone A., Fletcher C.F., Copeland N.G., Jenkins N.A., Crews S., Martinez S., Puellès L., Rubenstein J.L., Tessier-Lavigne M.;
RT Mol. Cell. Neurosci. 7:519-519(1996).
RN [3]
RP REVISIONS TO C-TERMINUS.
RX MEDLINE=97343329; PubMed=9199934;
RA Fan C.-M.;
RT Unpublished results, cited by:
RT Christ R., Scott H.S., Chen H., Kudoh J., Rossier C., Minoshima S., Wang Y., Shimizu N., Antonarakis S.E.;
RT Genome Res. 7:615-624(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=96413339; PubMed=8927054;
RA Ema M., Morita M., Ikawa S., Tanaka M., Matsuda Y., Gotoh O., Saijoh Y., Fujii H., Hamada H., Kikuchi Y., Fujii-Kuriyama Y.;
RT "Two new members of the murine Sim gene family are transcriptional repressors and show different expression patterns during mouse embryogenesis."
RT Mol. Cell. Biol. 16:5865-5875(1996).

[5]
RN SEQUENCE FROM N.A.
RP STRAIN-129/SV.
RA Hosoya T.;
RL Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.
RN [6]
RP SUBUNIT.
RX MEDLINE=97172525; PubMed=9020169;
RA Probst M.R., Fan C.M., Tessier-Lavigne M., Hankinson O.;
RT "Two murine homologs of the Drosophila single-minded protein that
interact with the mouse aryl hydrocarbon receptor nuclear
translocator protein";
RL J. Biol. Chem. 272:4451-4457(1997).
CC -1- FUNCTION: TRANSCRIPTIONAL FACTOR THAT MAY HAVE PLEIOTROPIC EFFECTS
DURING EMBRYOGENESIS AND IN THE ADULT.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
BHLH PROTEIN. HETERODIMER OF SIM1 AND ARNT.
CC -1- TISSUE SPECIFICITY: DETECTED IN LUNG, SKELETAL MUSCLE AND KIDNEY.
CC DURING FETAL DEVELOPMENT IT IS FOUND IN THE CNS, DEVELOPING
KIDNEY, MESONEPHRIC DUCT, AND FOREGUT.
CC SOMITES, MESONEPHRIC DUCT, AND FOREGUT.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
TRANSCRIPTION FACTORS.
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U40575; AAA91201.1; ALT_SEQ.
DR EMBL: D79209; BAA11467.1; -
DR EMBL: AB013491; BAA28270.1; -
DR EMBL: AB013484; BAA28270.1; JOINED.
DR EMBL: AB013485; BAA28270.1; JOINED.
DR EMBL: AB013486; BAA28270.1; JOINED.
DR EMBL: AB013487; BAA28270.1; JOINED.
DR EMBL: AB013488; BAA28270.1; JOINED.
DR EMBL: AB013489; BAA28270.1; JOINED.
DR EMBL: AB013490; BAA28270.1; JOINED.
DR MGD: MGI:98306; Sim1.
DR InterPro: IPR003015; HLH_MYC.
DR InterPro: IPR001092; HLH_dlm.
DR InterPro: IPR001067; Nucleinslocator.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00785; PAC; 1.
DR Pfam: PF00785; PAS; 2.
DR PRINTS: PR00785; NCTRNSLOCATR.
DR SMART: SM00353; HLH; 1.
DR SMART: SM00086; PAC; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
DR PROSITE: PS50112; PAS; 2.
KW Developmental protein; Neurogenesis; Nuclear protein; Repeat;
KW Transcription regulation; DNA-binding.
FT DNA_BIND 1 13 BASIC DOMAIN.
FT DOMAIN 14 34 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 77 147 PAS 1.
FT DOMAIN 218 288 PAS 2.
FT CONFLICT 133 133 H -> L (IN REF. 1).
FT CONFLICT 176 176 MISSING (IN REF. 1).
FT CONFLICT 322 322 P -> R (IN REF. 1).
FT CONFLICT 480 480 A -> P (IN REF. 1).
FT CONFLICT 537 537 D -> S (IN REF. 1).
SQ SEQUENCE 765 AA; 85540 MW; B1A7F7DA8578CD17 CRC64;

Query Match

82.1%; Score 32; DB 1; Length 765;

Best Local Similarity 85.7%; Pred. No. 47;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EVPGSOH 7
DB 423 EVPGSOH 429
RESULT 8
YH96_MYCLE
ID YH96_MYCLE STANDARD; PRT; 137 AA.
AC P13733;
DT 01-JAN-1990 (Rel. 13, Created)
DR 01-JAN-1990 (Rel. 13, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 15.2 kDa protein ML1796.
ML1796.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86808878; PubMed=2447183;
RA Booth R.J., Harris D.P., Love J.M., Watson J.D.;
RT "Antigenic proteins of Mycobacterium leprae. Complete sequence of the
gene for the 18-kDa protein";
RL J. Immunol. 140:597-601(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-TN.
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Elgimeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holtroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RL "Massive gene decay in the leprosy bacillus";
RL Nature 409:1007-1011(2001).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19058; AAA88230.1; -
DR EMBL: AL583923; CAC30749.1; -
DR PIR: B27386; B27586.
DR Leprona; ML1796; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 137 AA; 15195 MW; 633E68BF3FDAD0BB CRC64;

Query Match 79.5%; Score 31; DB 1; Length 137;
Best Local Similarity 83.3%; Pred. No. 12;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 VPGSOH 7
DB 127 VPGSOH 132
RESULT 9
CIDA_HUMAN
ID CIDA_HUMAN STANDARD; PRT; 219 AA.
AC O60543;

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cell death activator CIDF-A (Cell death-inducing DFFA-like effector
 A).
 GN CIDF-A.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=98232498; PubMed=9564035;
 RA Itohara N., Koseki T., Chen S., Wu X., Nunez G.;
 RT "cidf, a novel family of cell death activators with homology to the 45
 RT kha subunit of the DNA fragmentation factor.";
 RL EMBO J. 17:2526-2533(1998).
 CC -1- FUNCTION: ACTIVATES APOPTOSIS.
 CC -1- SUBUNIT: INHIBITED BY DFFB.
 CC -1- SIMILARITY: CONTAINS 1 CIDF-N DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF041378; AAC34987.1; -
 CC DR HSSP; 09UHD4; 1D4B.
 DR DR MIM; 604440; -
 DR DR InterPro; IPR003508; CAD.
 DR DR Pfam; PF02017; CIDF-N; 1.
 DR DR SMART; SM00266; CAD; 1.
 KW Apoptosis.
 FT DOMAIN 33 110 CIDF-N.
 FT SEQUENCE 219 AA; 24686 MW; 05F704823CE71C0E CRC64;
 SQ
 Query Match 79.5%; Score 31; DB 1; Length 219;
 Best Local Similarity 83.3%; Pred. No. 20;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 2 VPGSQH 7
 DB 109 MPGSQH 114
 RESULT 10
 ID YB5A_THEME STANDARD; PRT; 240 AA.
 AC P58009;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein TM1158.1.
 GN TM1158.1.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogales; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Unterbach T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser J.C.,
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 RT genome sequence of Thermotoga maritima.";

RL Nature 399:323-329(1999).
 RN [2]
 RP IDENTIFICATION.
 RA Medigue C., Bocs S.;
 RL Unpublished Observations (APR-2001).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE001773; -; NOT_ANNOTATED_CDS.
 DR DR TIGR; TM1158.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 240 AA; 27773 MW; 1BEF66C1C8BD2700 CRC64;
 QY 2 VPGSQH 7
 DB 226 VPGSQH 231
 RESULT 11
 ID DLX3_NOTVI STANDARD; PRT; 273 AA.
 AC P53770;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Homeobox protein DLX-3 (Box-4) (NVHBOX-4).
 GN DLX3 OR BOX4.
 OS Notopthalmus viridescens (Eastern newt) (Triturus viridescens).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Caudata; Salamandridae; Salamandridae;
 OC Notopthalmus.
 OX NCBI_TaxID=8316;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skin;
 RX MEDLINE=93050784; PubMed=1358728;
 RA Beauchemin M., Savard P.;
 RT "Two distal-less related homeobox-containing genes expressed in
 RT regeneration blastemas of the newt.";
 RL Dev. Biol. 154:55-65(1992).
 CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN REGENERATION BLASTEMAS. DETECTED
 CC IN FORELIMBS, HINDLIMBS, THE TAIL, FLANK, AND BRAIN AS WELL AS IN
 CC LIMB AND TAIL BLASTEMAS.
 CC -1- SIMILARITY: BELONGS TO THE DISTAL-LESS FAMILY OF HOMEBOX
 CC PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X63531; CAA45094.1; -
 DR HSSP; P22808; INK3.
 DR InterPro; IPR000047; HTH_repressr.
 DR DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF00046; homeobox; 1.
 DR PRINTS; PR00024; HOMEBOX.
 DR PRINTS; PR00031; HTHREPRESSR.

DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS00071; HOMEBOX_2; 1.
KW Homeobox: DNA-binding; Developmental protein; Nuclear protein.
FT DNA_BIND 125 184 HOMEBOX.
SQ SEQUENCE 273 AA; 30654 MW; B356D01233061F2F CRC64;

Query Match 79.5%; Score 31; DB 1; Length 273;
Best Local Similarity 71.4%; Pred. No. 26;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYVGSQH 7
||||:|
Db 188 EYVGMGH 194

RESULT 12

ID DLX3_PLEWA STANDARD; PRT; 274 AA.

AC Q91284;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Distal-less like protein DLX-3.

DLX3.

OS Pleurodeles waltl (Iberian ribbed newt).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae;

OC Pleurodeles.

OX NCBI_TaxID=8319;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-96391186; PubMed-8798159;

RA Nicolas S., Massacrier A., Caubit X., Cau P., le Parco Y.;

RT "A distal-less-like gene is induced in the regenerating central

RT nervous system of the urodele Pleurodeles waltl.";

RL Mech. Dev. 56:209-220(1996).

CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).

CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE REGENERATING SPINAL CORD BUT

CC NOT IN THE ADULT ONE.

CC -1- SIMILARITY: BELONGS TO THE DISTAL-LESS FAMILY OF HOMEBOX

CC PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Db 189 EYVGMGH 195

RESULT 13

ID DLX3_AMBME STANDARD; PRT; 280 AA.

AC Q90229;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Homeobox protein DLX-3.

DLX-3.

OS Ambystoma mexicanum (Axolotl).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Caudata; Salamandroidea; Ambystomatidae;

OC Ambystoma.

OX NCBI_TaxID=8296;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-97108743; PubMed-8951064;

RA Mullen L.M., Bryant S.V., Torok M.A., Blumberg B., Gardiner D.M.;

RT "Nerve dependency of regeneration: the role of Distal-less and FGF

RT signaling in amphibian limb regeneration.";

RL Development 122:3487-3497(1996).

CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).

CC -1- SIMILARITY: BELONGS TO THE DISTAL-LESS FAMILY OF HOMEBOX

CC PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Query Match 79.5%; Score 31; DB 1; Length 280;

Best Local Similarity 71.4%; Pred. No. 26;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYVGSQH 7
||||:|
Db 189 EYVGMGH 195

RESULT 14

ID COI2_BOVIN STANDARD; PRT; 414 AA.

AC Q9TR7;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE COUP transcription factor 2 (COUP-TF2) (COUP-TF II).

GN NR2P2 OR TFCOUP2.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Bovinae; Bos.

OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Corpus luteum;
 RL Walthers N.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REGULATION OF THE APOLIPOPROTEIN AI GENE TRANSCRIPTION.
 CC BINDS TO DNA SITE A (BY SIMILARITY).
 CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 CC NR2 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ249441; CAB5624.1; -.
 CC DR HSSP: P19793; 2NLL.
 CC DR InterPro: IPR000536; Hormone_rec_119.
 CC DR InterPro: IPR001723; Steroidhormone_receptor.
 CC DR InterPro: IPR001628; zf-c4.
 CC DR Pfam: PF00104; hormone_rec; 1.
 CC DR Pfam: PF00105; zf-c4; 1.
 CC DR PRINTS: PR00398; STRODHOMNER.
 CC DR PRINTS: PR00047; STROIDFINGER.
 CC DR SMART: SM00430; HOLI; 1.
 CC DR SMART: SM00399; ZNF_C4; 1.
 CC DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
 CC DR Transcription regulation; DNA-binding; Nuclear protein;
 CC Zinc-finger.
 CC KW
 CC FT DOMAIN 71 75 POLY-GLN.
 CC FT DNA_BIND 79 144 NUCLEAR RECEPTOR-TYPE.
 CC FT ZN_FING 79 99 C4-TYPE.
 CC FT ZN_FING 115 139 C4-TYPE.
 CC FT DOMAIN 337 414 IMPORTANT FOR DIMERIZATION (BY
 CC SIMILARITY).
 CC SO SEQUENCE 414 AA; 4557 MW; C24CB023C8A27F57 CRC64;
 Query Match 79.5%; Score 31; DB 1; Length 414;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 EVPGSQ 6
 Db 14 EVPGSQ 19
 RESULT 15
 COT2_HUMAN
 ID COT2_HUMAN STANDARD; PRT; 414 AA.
 AC P24468; Q03754;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE COUP transcription factor 2 (COUP-TF2) (COUP-TF II) (Apolipoprotein AI
 DE regulatory protein-1) (ARP-1).
 GN NR2F2 OR TRCOUP2 OR ARP1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91118002; PubMed=1899293;
 RA Ladias J.A.A., Karathanasis S.K.;
 RT "Regulation of the apolipoprotein AI gene by ARP-1, a novel member of
 the steroid receptor superfamily."

RL Science 251:561-565(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Speckmayer R.W.M., Paulweber B., Sandhofer F.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-351 FROM N.A.
 RX MEDLINE=92314709; PubMed=1820218;
 RA Wang L.H., Ing N.H., Tsai S.Y., O'Malley B.W., Tsai M.J.;
 RT "The COUP-TFs compose a family of functionally related transcription
 RT factors."
 RL Gene Expr. 1:207-216(1991).
 CC -1- FUNCTION: REGULATION OF THE APOLIPOPROTEIN AI GENE TRANSCRIPTION.
 CC BINDS TO DNA SITE A.
 CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: FOUND IN ALL TISSUES.
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 CC NR2 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M64497; AAA86429.1; -.
 CC DR EMBL: U60477; AAB09475.1; -.
 CC DR EMBL: M62760; AAA21479.1; -.
 CC DR PIR: A37133; A37133.
 CC DR HSSP: P19793; 2NLL.
 CC DR TRANSFAC: T00045; -.
 CC DR MIM: 107773; -.
 CC DR InterPro: IPR000536; Hormone_rec_119.
 CC DR InterPro: IPR001723; Steroidhormone_receptor.
 CC DR InterPro: IPR001628; zf-c4.
 CC DR Pfam: PF00104; hormone_rec; 1.
 CC DR Pfam: PF00105; zf-c4; 1.
 CC DR PRINTS: PR00398; STRODHOMNER.
 CC DR PRINTS: PR00047; STROIDFINGER.
 CC DR SMART: SM00430; HOLI; 1.
 CC DR SMART: SM00399; ZNF_C4; 1.
 CC DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
 CC DR Transcription regulation; DNA-binding; Nuclear protein;
 CC Zinc-finger.
 CC KW
 CC FT DOMAIN 71 75 POLY-GLN.
 CC FT DNA_BIND 79 144 NUCLEAR RECEPTOR-TYPE.
 CC FT ZN_FING 79 99 C4-TYPE.
 CC FT ZN_FING 115 139 C4-TYPE.
 CC FT DOMAIN 337 414 IMPORTANT FOR DIMERIZATION.
 CC SO SEQUENCE 414 AA; 4557 MW; C24CB2E8C8A27E8C CRC64;
 Query Match 79.5%; Score 31; DB 1; Length 414;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 EVPGSQ 6
 Db 14 EVPGSQ 19

Search completed: July 3, 2002, 09:04:41
 Job time: 581 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 08:55:46 ; Search time 74.17 Seconds

(Without alignments)
16.327 Million cell updates/sec

Title: US-09-786-648-2
Perfect score: 39
Sequence: 1 EPPGSH 7

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL.19:*
1: sp._archaea:*
2: sp._bacteria:*
3: sp._fungi:*
4: sp._human:*
5: sp._invertebrate:*
6: sp._mammal:*
7: sp._mhc:*
8: sp._organelle:*
9: sp._phage:*
10: sp._plant:*
11: sp._rodent:*
12: sp._virus:*
13: sp._vertebrate:*
14: sp._unclassified:*
15: sp._virus:*
16: sp._bacteriap:*
17: sp._archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	39	100.0	103	2	Q9R646
2	39	100.0	124	2	Q57193
3	39	100.0	124	2	Q9R615
4	39	100.0	124	2	Q56635
5	39	100.0	124	2	Q93V32
6	39	100.0	124	9	Q94M01
7	35	89.7	565	10	Q22511
8	34	87.2	195	13	Q9W7D3
9	34	87.2	199	13	Q9W7D4
10	34	87.2	201	2	Q9R3Z8
11	33	84.6	192	4	Q9NP06
12	33	84.6	272	4	Q9UG04
13	33	84.6	367	10	Q80418
14	33	84.6	412	4	Q9Y2B3
15	33	84.6	868	5	Q9VFE2
16	32	82.1	95	10	Q9F018

17	32	82.1	137	16	Q92T78	Q92T78 rhizobium m
18	32	82.1	765	11	Q70284	Q70284 mus musculu
19	31	79.5	95	10	Q9F0J8	Q9F0J8 amarantinus
20	31	79.5	122	5	Q9W2V0	Q9W2V0 drosophila
21	31	79.5	134	16	Q92N40	Q92N40 rhizobium m
22	31	79.5	222	11	Q9JL95	Q9JL95 mus musculu
23	31	79.5	233	10	Q94KK7	Q94KK7 arabidopsis
24	31	79.5	242	2	Q86582	Q86582 streptomyce
25	31	79.5	243	11	Q9QY63	Q9QY63 mus musculu
26	31	79.5	260	10	Q9MA16	Q9MA16 arabidopsis
27	31	79.5	274	17	Q9HNM5	Q9HNM5 halobacteri
28	31	79.5	331	5	Q18391	Q18391 drosophila
29	31	79.5	353	16	Q9RVJ7	Q9RVJ7 delnoccocus
30	31	79.5	360	11	Q9D6W8	Q9D6W8 mus musculu
31	31	79.5	371	2	Q9RH15	Q9RH15 zymomonas m
32	31	79.5	374	16	Q9PKY2	Q9PKY2 chlamydia m
33	31	79.5	382	5	Q9V7M7	Q9V7M7 drosophila
34	31	79.5	406	3	Q96V07	Q96V07 cryphonectr
35	31	79.5	432	3	Q9P3J3	Q9P3J3 neurospora
36	31	79.5	459	11	Q9R0C1	Q9R0C1 mus musculu
37	31	79.5	460	10	Q9LIF8	Q9LIF8 arabidopsis
38	31	79.5	472	2	Q939I3	Q939I3 klebsiella
39	31	79.5	489	5	Q9XUM8	Q9XUM8 caenorhabdi
40	31	79.5	496	11	Q9D0H2	Q9D0H2 mus musculu
41	31	79.5	497	4	Q9BRR6	Q9BRR6 homo sapien
42	31	79.5	510	4	Q9H846	Q9H846 homo sapien
43	31	79.5	510	4	Q96GM8	Q96GM8 homo sapien
44	31	79.5	559	6	Q95JP7	Q95JP7 macaca fasc
45	31	79.5	660	10	Q9AY34	Q9AY34 oryza sativ

ALIGNMENTS

RESULT 1
ID Q9R646 PRELIMINARY; PRT: 103 AA.
AC Q9R646;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CHOLERA-LIKE ENTEROTOXIN B SUBUNIT.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
NC NCB1_TaxID=666;
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE.
RX MEDLINE=95303036; PubMed=7783690;
RA Nakashima K., Eguchi Y., Nakasone N.;
RT "Characterization of an enterotoxin produced by Vibrio cholerae
RT O139.";
RL Microbiol. Immunol. 39:87-94(1995).
DR HSSP: P01556; IYTC.
DR InterPro: IPR001835; Enterotoxin_B.
DR Pfam: PF01376; Enterotoxin_B.1.
DR PRINTS: PR00772; ENTEROTOXINB.
DR PRODOM: PD012805; Enterotoxin_B.1.
SQ SEQUENCE 103 AA; 11645 MW; 992A05C07F2F70E CRC64;

Query Match 100.0%; Score 39; DB 2; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPPGSH 7
Db 51 EPPGSH 57
RESULT 2
ID Q57193 PRELIMINARY; PRT: 124 AA.
AC Q57193;

DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE CHOLERA TOXIN B PROTEIN (CTB) PRECURSOR (CTB).
GN CTXB.
OS *Vibrio cholerae*.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; *Vibrio*.
OX NCBI_TaxId=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CLASSICAL STRAIN 569B;
RX MEDLINE=91355224; PubMed=1683840;
RA Dams E., De Wolf M., Dietrick W.;
RT "Nucleotide sequence analysis of the CT operon of the *Vibrio cholerae*
RL classical strain 569B.";
RL Biochim. Biophys. Acta 1090:139-141(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CLASSICAL BIOTYPE 569B;
RA Shi C., Cao C., Zhang J., Ma Q.;
RL Chin. Biochem. J. 9:395-399(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-CLASSICAL BIOTYPE 569B;
RA Xu L.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X58785; CAA41591.1; -;
DR EMBL; U25679; AAC34728.1; -;
DR EMBL; A00931; CAA00098.1; -;
DR HSSP; P01556; 2CHB.
DR InterPro: IPR001835; Enterotoxin_B.
DR Pfam: PF01376; Enterotoxin_B; 1.
DR PRINTS; PR00772; ENTEROTOXINB.
DR ProDom: PD012805; Enterotoxin_B; 1.
KW signal.
FT SIGNAL.
FT CHAIN 1 21 POTENTIAL.
FT SEQUENCE 124 AA; 13919 MW; D6BF83FF7924EA3 CRC64;
SQ

Query Match 100.0%; Score 39; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
Db 72 EVPGSOH 78

RESULT 3
Q9RP15 PRELIMINARY; PRT; 124 AA.
AC Q9RP15;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE CHOLERA ENTEROTOXIN B-SUBUNIT.
GN CTXB.
OS *Vibrio cholerae*.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; *Vibrio*.
OX NCBI_TaxId=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KNIH002;
RA Shin H.J., Park Y.C., Kim Y.C.;
RT "Cloning and nucleotide sequence analysis of the virulence gene
RL cassette from *Vibrio cholerae* KNIH002 isolated in Korea.";
RL Misalimurnag Hoji 35:205-210(1999).
DR EMBL; AF175708; AAD51360.1; -;
DR HSSP; P01556; 2CHB.
DR InterPro: IPR001835; Enterotoxin_B.
DR Pfam: PF01376; Enterotoxin_B; 1.
DR PRINTS; PR00772; ENTEROTOXINB.
RN

DR ProDom: PD012805; Enterotoxin_B; 1.
SQ SEQUENCE 124 AA; 13905 MW; 23BF83FF7935B9 CRC64;

Query Match 100.0%; Score 39; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
Db 72 EVPGSOH 78

RESULT 4
Q56635 PRELIMINARY; PRT; 124 AA.
AC Q56635;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE CHOLERA TOXIN PRECURSOR.
GN CTXB.
OS *Vibrio cholerae*.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; *Vibrio*.
OX NCBI_TaxId=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S7;
RA Yamamoto K., Do V.G., Xu M., Iida T., Miwatani T., Albert M.J.,
RA Honda T.;
RT "Comparison of cholera toxin genes (ctxAB) of non-O1 *Vibrio cholerae*
RL strains 854 (O139-bengal) and S7 (O37) from two outbreaks.";
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; D30052; BAA06289.1; -;
DR HSSP; P01556; 2CHB.
DR InterPro: IPR001835; Enterotoxin_B.
DR Pfam: PF01376; Enterotoxin_B; 1.
DR PRINTS; PR00772; ENTEROTOXINB.
DR ProDom: PD012805; Enterotoxin_B; 1.
KW signal.
FT SIGNAL.
FT CHAIN 1 21 POTENTIAL.
FT SEQUENCE 124 AA; 13871 MW; 3F87B2F297953179 CRC64;
SQ

Query Match 100.0%; Score 39; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
Db 72 EVPGSOH 78

RESULT 5
Q93V32 PRELIMINARY; PRT; 124 AA.
AC Q93V32;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HEAT-LABILE ENTEROTOXIN B SUBUNIT.
GN LTH B SUBUNIT.
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxId=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-1032 (ENTEROTOXIGENIC);
RA Komase K.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

RC STRAIN-1032 (ENTEROTOXIGENIC);
 RX MEDLINE-95091056; PubMed-7998417;
 RA Tamura S., Asanuma H., Tomita T., Komase K., Kawahara K., Danbara H.,
 RA Hattori N., Watanabe K., Suzuki Y., Nagamine T., Aizawa C., Oya A.,
 RA Kurita T.;
 RT "Escherichia coli heat-labile enterotoxin B subunits supplemented with
 RT a trace amount of the holotoxin as an adjuvant for nasal influenza
 RT vaccine.";
 RL Vaccine 12:1083-1089(1994).
 DR EMBL; AB011677; BAA25726.1; -
 SQ SEQUENCE 124 AA; 14028 MW; 5346BD38B32354C2 CRC64;

Query Match 100.0%; Score 39; DB 2; Length 124;
 Best Local Similarity 100.0%; Pred. No. 0.65;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7
 DB 72 EYVGSQH 78

RESULT 6
 O94M01 PRELIMINARY; PRT; 124 AA.
 AC O94M01:
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CTXB.
 GN CTXB.
 OS Vibrio phage CTX.
 OC Viruses; ssDNA viruses; Inoviridae; Inovirus.
 OX NCBI_TaxID-141904;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bhattacharyya T., Nandy R.K., Nair G.B.;
 RT "The entire core region of the ctx-phi (ctx-phage) in VCE 232, an
 RT environmental strain of V. cholerae.";
 RL Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF414369; AAL09682.1; -
 SQ SEQUENCE 124 AA; 13884 MW; 8D4C251591B16891 CRC64;

Query Match 100.0%; Score 39; DB 9; Length 124;
 Best Local Similarity 100.0%; Pred. No. 0.65;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7
 DB 72 EYVGSQH 78

RESULT 7
 O22511 PRELIMINARY; PRT; 565 AA.
 ID O22511:
 AC O22511:
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE GLUTATHIONE REDUCTASE (NADPH) (EC 1.6.4.2) (FRAGMENT).
 GN GOR.
 OS Vitis vinifera (Grape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae;
 OC Vitis.
 OX NCBI_TaxID-29760;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. THOMPSON SEEDLESS (CLONE 2A); TISSUE=FRUIT;
 RA Casool T., Adams D.O.;
 RT "Cloning of a grape glutathione reductase cDNA and analysis of its
 RT expression during berry development.";

RL Submitted (Aug-1997) to the EMBL/Genbank/DBJ databases.
 CC -I- COFACTOR: FAD (BY SIMILARITY).
 DR EMBL; AF019907; AAB70837.1; -
 DR HSSP; P06715; 1GER.
 DR InterPro: IPR001327; FAD_pyr_redox.
 DR InterPro: IPR000103; Pyridine_redox_2.
 DR InterPro: IPR001100; pyr_redox.
 DR InterPro: IPR004099; pyr_redox_dim.
 DR Pfam: PF00070; pyr_redox_1.
 DR Pfam: PF02852; pyr_redox_dim; 1.
 DR PRINTS: PR00368; FADPNR.
 DR PRINTS: PR00411; PNDRFASFI.
 DR PRINTS: PR00469; PNDRFASFI.
 DR PROSITE: PS00076; PYRIDINE_REDOX_1; 1.
 DR FAD; Flavoprotein; Oxidoreductase.
 FT NON_TER
 SQ SEQUENCE 565 AA; 60695 MW; B26113AE09A121DE CRC64;

Query Match 89.7%; Score 35; DB 10; Length 565;
 Best Local Similarity 71.4%; Pred. No. 25;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7
 DB 239 EYVGSQH 245

RESULT 8
 O9W7D3 PRELIMINARY; PRT; 195 AA.
 ID O9W7D3:
 AC O9W7D3:
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL 21.4 KDA PROTEIN.
 OC Oryzias latipes (Medaka fish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthopterygii; Acanthopterygii; Percomorpha; Atherinomorpha;
 OC Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
 OX NCBI_TaxID-8090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ORANGE-RED;
 RX MEDLINE-20070874; PubMed-10602271;
 RA Kanamori A.;
 RT "Systematic identification of genes expressed during early oogenesis
 RT in medaka.";
 RL Mol. Reprod. Dev. 55:31-36(2000).
 DR EMBL; AF12818; AAD38915.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 195 AA; 21446 MW; B90682D695729E88 CRC64;

Query Match 87.2%; Score 34; DB 13; Length 195;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 VPSQH 7
 DB 111 VPSQH 116

RESULT 9
 O9W7D4 PRELIMINARY; PRT; 199 AA.
 ID O9W7D4:
 AC O9W7D4:
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL 21.7 KDA PROTEIN.
 OS Oryzias latipes (Medaka fish).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthopterygii; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.
OX NCBI_TaxID=8090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORANGE-RED;
RX MEDLINE=20070874; PubMed=10602271;
RA Kanamori A.;
RT "Systematic identification of genes expressed during early oogenesis
in medaka.";
RL MOJ. Reprod. Dev. 55:31-36(2000).
DR EMBL; AF128817; AAD38914.1; -.
KW Hypothetical protein.
SQ SEQUENCE 199 AA; 21726 MW; E9827C05451B15CD CRC64;

Query Match 87.2%; Score 34; DB 13; Length 199;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 EVPGSOH 7
Db 115 EVPGSOH 120

RESULT 10

O9RJZ8 PRELIMINARY; PRT; 201 AA.

AC O9RJZ8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PUTATIVE AMIDASE.
GN SCG11A.03.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kiese R.M., Denapite D., Eichner A., Cullum J.,
RA Khashi H., Hopwood D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL MOJ. Microbiol. 21:77-96(1996).
DR EMBL; AL133210; CAB61584.1; -.
DR InterPro: IPR002502; Amidase_2.
DR Pfam: PF01510; Amidase_2; 1.
SQ SEQUENCE 201 AA; 22749 MW; B8F477E06A20468 CRC64;

Query Match 87.2%; Score 34; DB 2; Length 201;
Best Local Similarity 85.7%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
Db 176 EVPGSOH 182

RESULT 11
O9NPQ6 PRELIMINARY; PRT; 192 AA.

ID O9NPQ6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE LPL, LCAT-LIKE LYOPHOSPHOLIPASE (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pluvinet R., Estivill X., Escarceller M., Sumoy L.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Aulifay C., Ansoerge W., Ballabio A., Estivill X., Gibson K.,
RA Lehrach H., Poustka A., Lundeberg J.;
RT "The European IMAGE consortium for integrated Molecular analysis of
human gene transcripts.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL389957; CAB97531.1; -.
FT NON_TER 1
SQ SEQUENCE 192 AA; 21609 MW; 04A7AE8CB344F213 CRC64;

Query Match 84.6%; Score 33; DB 4; Length 192;
Best Local Similarity 71.4%; Pred. No. 21;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
Db 166 EVPGSOH 172

RESULT 12
O9UG04 PRELIMINARY; PRT; 272 AA.

AC O9UG04;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHETICAL 31.0 KDA PROTEIN (UNKNOWN) (PROTEIN FOR MGC:12657).
GN DKF2P564A0122.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Wambolt R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN; NEUROBLASTOMA;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL110209; CAB53675.1; -.
DR EMBL; BC011640; AAH1640.1; -.
DR InterPro: IPR003386; LACT.
DR Pfam: PF02450; LACT; 1.
KW Hypothetical protein.
SQ SEQUENCE 272 AA; 31016 MW; ACCCE5E1680D7A720 CRC64;

Query Match 84.6%; Score 33; DB 4; Length 272;
Best Local Similarity 71.4%; Pred. No. 31;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
|:|:|:|
Db 246 ELPGSEH 252

RESULT 13
ID 080418 PRELIMINARY; PRT; 367 AA.
AC 080418;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NTC16 PROTEIN.
GN NTC16.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RA Tanaka-Ideguchi M., Itoh H., Oyama N., Koshioka M., Matsuo M.;
RT "Over-expression of a tobacco homeobox gene, NTH15, decreases the
RT expression of a gibberellin biosynthetic gene encoding GA 20-
RT oxidase.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB016084; BAA31690.1; -
DR InterPro: IPR002419; Fe_Asc_oxidored. 1.
DR Pfam: PF00671; Fe_Asc_oxidored; 1.
SQ SEQUENCE 367 AA; 42170 MW; 923BC90B3BBAC05 CRC64;

Query Match 84.6%; Score 33; DB 10; Length 367;
Best Local Similarity 85.7%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
|:|:|:|
Db 149 EVPGSOH 155

RESULT 14
ID 09Y2B3 PRELIMINARY; PRT; 412 AA.
AC 09Y2B3;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE LCAT-LIKE LYSOPHOSPHOLIPASE (LPL).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99194552; PubMed=10092508;
RA Tanigawa Y., Shibata S., Kita S., Horikoshi K., Shirafuji H.,
RA Shimizu Y., Fujino M.;
RT "Cloning and expression of a novel lysophospholipase which
RT structurally resembles lecithin cholesterol acyltransferase.";
RL Biochem. Biophys. Res. Commun. 257:50-56(1999).
DR EMBL: AB017494; BAA76877.1; -
DR InterPro: IPR003386; LACT.
DR Pfam: PF02450; LACT; 1.
SQ SEQUENCE 412 AA; 46657 MW; 1FEA8A5783AF050A CRC64;

Query Match 84.6%; Score 33; DB 4; Length 412;
Best Local Similarity 71.4%; Pred. No. 49;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
|:|:|:|
Db 1 EVPGSOH 7

Db 386 ELPGSEH 392

RESULT 15
ID 09VEZ2 PRELIMINARY; PRT; 868 AA.
AC 09VEZ2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CG3837 PROTEIN.
GN CG3837.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkov D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.E., Ferraz C., Ferreira S., Fleischmann W.,
RA Fouts D., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodde C., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Howack J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang X., Lin X.,
RA Liu X., Mates B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Moberg C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Sylvestre R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.W., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003706; AAF55118.1; -
DR FlyBase: FBgn0038279; CG3837.
DR InterPro: IPR000494; EGFR-L.
DR Pfam: PF01030; Recep_L_domain; 2.
SQ SEQUENCE 868 AA; 98349 MW; F6562A64E72E7B21 CRC64;

Query Match 84.6%; Score 33; DB 5; Length 868;
Best Local Similarity 71.4%; Pred. No. 11e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
|:|:|:|
Db 812 ELPCTOH 818

Mon Jul 8 07:51:46 2002

us-09-786-648-2.ispt

Page 6

Search completed: July 3, 2002, 09:06:02
Job time: 616 sec